PSYCHOLOGICAL ASPECTS OF SURVEY METHODOLOGY:

EXPERIMENTS ON THE RESPONSE PROCESS

Margaret Kathleen Murphy

Thesis submitted for the degree of

Doctor of Philosophy

of the University of London.

January 1996.

Department of Social Psychology,

London School of Economics and Political Science,

University of London

UMI Number: U616002

UMI U616002

ABSTRACT

This thesis examines the psychological processes involved in responding to survey questions. Minor variations in questions have been shown to lead to variation in responses. These findings are inconsistent with the assumption that survey questions are tapping stable responses. Recently, psychological theories have been used to provide an explanation for these response effects.

Research applying psychological theory to survey response is reviewed, covering research on both behavioural and attitudinal questions. These reviews illustrate a reconceptualisation of the basis of the survey response. The need for more detailed data on the response process is identified. Verbal reports are identified as a potential method for producing process data, yet, uncertainty over their validity is noted. The use of verbal reports as data is then reviewed, covering both their historical and more recent use.

In the present research verbal report techniques are first experimentally examined to find an appropriate technique for obtaining process data in surveys. Think-aloud techniques are then used to examine the processes involved in responding to questions. A split-ballot questionnaire was administered, varying a number of questionnaire features where response effects have been hypothesised or shown to occur. Generally, the verbal protocols showed processing differences between the

different question forms, and provided information about general types of cognitive processing during response.

The next study moved on to look at context effects for attitudinal questions. An experiment was carried out in which a number of factors hypothesised to be influential in producing context effects were examined. A questionnaire was administered via computer and response latencies were measured. The results provide further information about the nature of context effects at attitude questions.

The findings from this study are then discussed in terms of the methodologies used, the specific response effects addressed, and the survey response process generally.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

9

11

LIST OF FIGURES

INTRODUCTION

This investigation concerns the application of psychological theory and methods to the understanding of survey methodology. Survey methods have a long history of use both as a research tool within the social sciences as well as for data collection in other areas. Research examining the reliablity of this method has shown that small variations to survey questions can lead to differences in response (Schuman & Presser, 1981). For example, small changes in question wording, in the type of response options provided or in the context in which the question is asked can produce differences in response. These findings are described in Chapter one.

Attempts to explain these findings have led to the application of psychological theory to survey methods, which has the broader aim of understanding of the processes which produce response to questions. Chapter one discusses assumptions and theories about the processes which produce responses to survey questions. Research within the CASM (Cognitive Aspects of Survey Methodology) framework is then reviewed. Survey questions include questions about behaviour and attitudes. Chapter one reviews work within the CASM framework which looks at the processes involved in response to behavioural questions. Conceptions of attitudes are then discussed and research focusing on context effects for attitude questions is reviewed.

Attempts to explain the cognitive processes which occur during response to survey questions leads to the need to provide data

on these processes. To date most research has relied on indirect, response data on processing. In part this thesis aims to explore methods for providing information on response processes. Verbal protocols are identified as a potential method. Yet this method has been surrounded by controversy throughout its history of use in psychology. Chapter 2 explores the controversies surrounding the use of verbal protocols and indicates how the use of verbal reports applies to survey methods.

Chapter 3 reports on an experimental investigation of the use of different verbal report techniques in the survey. It aims to provide the basis for choosing a suitable method for use in surveys. Qualitative assessments are made of the techniques in terms of their validity as reports of ongoing processing. Assessments of the completeness of the verbal reports are also made. A methodology for think aloud protocols is developed from this experiment and is used in subsequent studies.

Chapters 4 and 5 report a study which uses think aloud reports with split ballot questionnaires. A number of response issues concerning both behavioural and attitudinal questions are addressed. In some cases hypotheses concerning particular questions were drawn from split-ballot surveys, in other cases untested questions which addressed particular issues were used. These included issues of the influence of response scales on response, assimilation and contrast effects, and the order in which questions are asked. The aim is not to explore one response effect in detail but to address a range of

issues.

Chapters 6 and 7 move on from verbal reports to a different method. Perhaps the major issue in response effects for attitudinal questions is context effects. Context effects can arise when previous questions influence the response to later questions. A response effect can be shown when different previous questions, or contexts, produce different responses to a particular question. Context effects may be in the direction of assimilation, producing responses consistent with the previous context, or they may be in the direction of contrast, producing responses contrary to the previous context. In part, these effects, and the direction of the effect, may depend on the amount of time and cognitive effort a respondent brings to the answering task. Think-aloud protocols were deemed inappropriate for exploring such a factor. Thus, a split ballot questionnaire was used with manipulations, and measurement, of response time. A number of factors which have been suggested to be influential in producing context effects were addressed.

Finally chapter 8 reviews the findings of the studies. The methods used in this study are reviewed as to their appropriateness for exploring response processes in the survey. The findings are discussed in terms of the specific types of response effects addressed and in terms of the response process in general. Suggestions are made as to what factors are likely to be important considerations in response effects, as well as a discussion of the survey more generally.

CHAPTER 1:   THE SURVEY METHOD

SUMMARY

The assumptions traditionally seen as underlying the survey method are discussed. Contradictory findings which undermine these assumptions are reviewed, illustrating a need to reconsider the basis of survey response. Explanations of the survey response in terms of psychological theory are discussed and research using psychological theory to explain survey response is reviewed. The need for a reconceptualisation of survey measurement is discussed in light of this review.

1.1 WHAT KIND OF METHOD IS THE SURVEY METHOD?

Asking questions and giving answers to questions is a fairly mundane activity; people do it all the time. Survey research has taken this basic activity and developed it into a research method. Essentially the survey elicits information from people by asking them questions, and this simple technique is one of the most widely used research tools. Surveys provide information about people's behaviours, attitudes, beliefs, states of mind etc. They provide information that is often used as the basis for policy, marketing, and other decisions, as well as providing an important research tool in the social sciences.

Central to the use of survey questioning is the belief that people can answer questions reliably and validly. That is,

the responses offered to the questions posed are meaningful. This belief essentially rests on two assumptions. The first assumption is quite aptly put by Fischhoff (1991) when, in discussing the measurement of values, he asks the question "is there anything in there?". He describes a continuum of philosophies on the expression of values from the philosophy of articulated values to that of basic values. The former assumes that people have answers to the questions they are asked whilst the latter assumes that usually they do not, but must rather derive responses from more basic values. Different positions along this continuum have led to different research paradigms. Fischhoff places survey researchers at the articulated values end of the continuum. That is survey researchers have faith that there is something 'in there'. A second, related assumption is that what is in there can, in principle, be 'got out'.

This belief by survey researchers that respondents can give meaningful answers concerns not only answers to value questions, but extends to other types of questions asked on surveys, including those addressing behaviours and attitudes. Given the assumption that there is something 'in there' to be 'got out' the focus in survey research has been how to 'get out' what is already 'in there'.

Brenner (1982) applies the stimulus - response analogy to survey response, whilst noting that survey response is more complex. In these terms the problem is to find the right stimulus to elicit the desired response. One central tenet is

that a questionnaire must be standardised; each respondent
must receive exactly the same question, or stimulus. Even
minor differences might mean that a different stimulus has
been administered and thus, as Fischhof puts it, "Any slip
could evoke a precise, thoughtful answer to the wrong
question" (1991, p839).

Essentially, problems encountered with survey results, for
example slightly different wordings leading to different
responses, were not treated as problems with respondents'
inability to produce a response, but as technical problems
with the questionnaire itself. Hyman for instance says: "when
the original conception is good, technical errors in
interviewing or in the construction of questionnaires can
damage the results, but these are technical problems only"
(1954, p666). Similarly, Lazarsfeld (1944), in defending the
merits of both depth interviewing and closed survey
questioning, was concerned with getting a picture of the
respondent's 'real' attitude.

Producing a good questionnaire was not assumed to be an easy
task. Many textbooks describing the techniques of survey
design were produced (eg Hyman, 1955; Payne, 1951). The focus
was on how to get questions, and interviews, right. For
example, methods for producing unbiased questions were
developed (Suchman & Guttman, 1947). These considerations
aimed to produce the best method of getting at what is there
without leading the respondent away from his or her true
response.

Not only questions but also interviewers, could bias results by, for example, suggesting answers to respondents. Work on interviewer effects is analogous to concerns in experimental psychology with the effects of experimenter bias (Rosenberg, 1969; Rosenthal & Rosnow, 1969). Interviewers are thus trained to read the questions exactly as presented and to exert no directive influence on the respondent.

However, there were concerns other than simply the correctness of the instrument and its administration. Whether the respondent was telling the truth was another concern (Hyman, 1944). Again this reflects the general idea that a response is there, but the concern is whether the respondent is willing to report it accurately. This underlying assumption of the ability of the respondent to answer, if they want to, is shown in a study of voting behaviour (Miller, 1952). Comparing reported behaviour to actual behaviour it was found that 22 people had 'lied'. That is they reported that they had voted when in fact they had not. There is no other explanation offered for this discrepancy (eg the respondent could not remember and so misreported). The assumption is that they could have reported accurately if they had wanted to, but for some reason they did not want to. Guarding against lying, and associated motivational biases, also became important to survey researchers. Much work examined the various kinds of response sets that affected people's responses to questions. Greenberg & Folger (1988) compare this work on response sets to that on subject roles in the experimental situation (Orne, 1962; Weber & Cook, 1972). Thus, as in experiments, one had

to protect respondents from undue pressures to respond in a particular way. Given the right question in a situation free from biasing factors, respondents could report what they thought or did relatively unproblematically.

In part this characterisation of survey research is simplified. Early researchers raised as a possibility the question of whether there was really anything there to be measured; if only in passing they suggested that people might answer attitude questions off the top of their head (eg Sanford, 1951). Later research on non-attitudes (Converse, 1974) made this question a more salient and central issue. Memory problems associated with asking retrospective questions were also recognised (eg Kendall & Lazarsfeld, 1954). However, although these issues were considered, they did not provoke a reconceptualisation of what was being measured. The implications of these issues were not fully pursued. The crucial issue of what if anything was 'in there' was largely side-lined.

However, research findings showing response effects were still found, even with good questions developed and administered by experienced survey practitioners. Much of the instigation for a rethink of survey methodological assumptions has come from the accumulated evidence of such response effects. Survey researchers found anomalous results and had no coherent explanation for them. I now want to look briefly at the types of response effects that have led to such a rethink,

before moving on to look at the more recent attempts to understand the survey response process.

1.1.2 Anomalies -- 'errors' in measurement

Bradburn (1985) distinguishes two major sources of errors in surveys: sampling and non-sampling errors. The latter he further divides into errors due to failures in executing the sample and errors due to other factors, which are better termed response effects. There is a good deal of knowledge about sampling error and errors associated with the sample, but much less is known about the possible biasing effects involved in responding.

Response effects have generally been conceived of as occurring during the interview and have been ascribed to three sources: the interviewer, the respondent and the task (Bradburn and Sudman, 1979). Interactions between these elements can also be a source of error. Research on effects due to the interviewer has included both extra-role characteristics, such as race (eg. Hagenaars and Heinen, 1982) and role-restricted characteristics such as misreading the question (eg. Brenner, 1982), or style of interviewing (Dijkstra, Van der Veen, & Van der Zouwen, 1985). Research on effects due to the respondent has also included demographic characteristics (eg. Gove, 1982) and other, more psychological, factors such as motivation and self-serving biases (eg. Phillips and Clancy, 1972).

Although there has been much research on these factors, with equivocal results, the task itself seems to be the most important source of response effects (Bradburn, 1985).

The task can include mode of administration, where, typically, research has been equivocal about the difference in response effects for different modes (Cannell, 1985a). One can also focus on the questionnaire as the task, and, indeed, much of the research on response effects due to the task concentrates on this area.

In general, the questionnaire involves three areas where response effects may occur. These are the question wording, the response alternatives, and the context of the question. Several comprehensive reviews of research on these areas exist (eg Turner and Martin, 1984; Schuman and Presser, 1981). Generally, changes, often minor changes, to any of these questionnaire features can sometimes have substantial effects on the distribution of responses.

Differences to the wording of a question can, for example, result in the questions tapping different aspects of a multifaceted issue (Turner and Martin, 1984). Very minor wording differences may also produce different results, for example, using either the word 'welfare' or the word 'poor' (Smith, 1987). A classic example of wording effects is a question asking either whether the US should **forbid** public speeches against democracy, or asking whether they should **not**

allow them. Different results were obtained with the different wordings (Rugg, 1941; Schuman and Presser, 1981).

Differences in results have also been obtained by changing the response categories offered. A number of changes in response categories have been shown to affect responses. For example, offering a middle category, and its position in the list of options (Bishop, 1987), and explicitly including or excluding a 'don't know' category (Schuman and Scott, 1989).

More subtle, yet equally substantial, effects have been obtained by altering the context in which a question is asked. A classic example of this is the 'communist reporter' question (Schuman and Presser, 1981; Schuman, Kalton, & Ludwig, 1983). If people are asked first whether a communist country should allow American reporters to send news to America, they are then more likely to approve of allowing communist reporters to send news from America than if the order of questions is reversed.

The kinds of effects described above have led survey researchers to reconsider the assumptions on which survey methods are based. As Schuman et al (1983) say they have been "forced to regard the problem of context as a matter of real substantive importance rather than a technical issue". The faith in technical solutions had been shaken; the search for alternate, coherent explanations was instigated.

Although there were always attempts to explain the various response effects that occur (for example, Schuman and Presser (1981) explain the context effects produced by the communist reporter questions by suggesting that it evokes a norm of reciprocity) these explanations have tended to be fragmentary and lacking a coherent theoretical basis. One of the many recommendations made by Turner and Martin (1984) for improving our understanding of the survey is the need for theoretically based research. This research is needed in order to understand why effects that have been demonstrated occur, but also to guide the selection of factors that may be related to response effects. Given that the assumptions underlying surveys have been shown to be questionable, what should we assume underlies survey response?

1.2 COGNITIVE ASPECTS OF SURVEY METHODOLOGY --

A NEW THEORY OF SURVEY RESPONSE?

Two conceptualisations of the survey have emerged, which approach the problem from different perspectives, and, perhaps, at different levels. On the one hand, coming from more sociological perspectives are conceptualisations regarding the interview as a communicative event (Briggs, 1986; Mishler, 1986; Suchman & Jordan, 1992). These approaches concentrate on the interview, rather than for example self-completion questionnaires. Briggs presents a model of the interview that locates it within a social context, involving types of communicative events and social situations, as well as the goals of participants and their

roles. He argues that interviews largely involve indexical meaning, that is, meaning tied to a context (for the research interview the context is previous questions as well as the surrounding situation). However, the standard survey approach is to treat interviews as though they involve referential meaning (ie. meaning not tied to a particular context). Neglecting the contextual nature of interaction distorts the interpretation placed on responses. Mishler (1986), similarly, suggests that a survey interview involves the joint construction of meaning by participants. He argues that the practices of survey interviewing (standardisation, limited interviewer involvement, etc.) deny this construction of meaning and undermine the validity of responses.

In this view there is not necessarily a conceptualisation of what respondents intrinsically can and cannot report, but rather the focus is upon the problems of the interview as a form through which views are expressed. The survey interview, as currently conducted, constitutes a barrier to valid reporting. The suggestion is that survey interview practices must be changed away from standardised practices to practices that allow exchanges between respondent and interviewer to determine joint understanding. The focus is on social interaction, with an emphasis on language. Whilst I think this conceptualisation has something to offer in explaining the problems inherent in this interaction, it is also limited in scope and thus limited in the problems it can address. It tends to reach neither beyond itself to explore the social structures that may shape response to various issues nor

within itself to explore the variety of individual factors that shape response.

On the other hand, drawing on cognitive and social psychology, is a conceptualisation of the survey largely in terms of the cognitive processes underlying responses. This approach has received far more research attention than the approach described above, and it is this approach that will be discussed further. The application of psychology to survey practice was in part a direct effort by survey researchers to deal with the problems of response effects, but also it was seen by psychologists as a way of exploiting surveys for cognitive psychological research (Loftus, Fienberg and Tanur, 1985). It was thus through the joint efforts of survey researchers and psychologists that this research programme developed (Jabine, Straf, Tanur, & Tourangeau, 1984; Jobe & Mingay, 1991). The outcome of this effort has been the development of research that attempts to provide a general theoretical framework for the explanation of response behaviour in the survey in terms of psychological processes.

It might be suggested that early survey researchers' conceptualisation of response processes lacked sophistication. However, it might also be claimed that the psychological theories extant at the time may not have been up to the task of offering a much more sophisticated conceptualisation. The present approach draws heavily on recent work in cognition and especially social cognition. This area adopts, at least by analogy, an information processing approach to cognition

(Hamilton, Devine & Ostrom, 1994). Generally, the conception of the cognitive system is a dynamic one, where various aspects of the cognitive system interact (Wyer & Srull, 1986). The information drawn on can be abstract or particular, (eg schemas, or MOPS, Schank, 1987; exemplars Kahneman & Miller, 1986). The way information is structured is also considered. Distinctions have been made between different types of processing, for example controlled or automatic (Schneider & Shiffrin, 1977) , deliberate or spontaneous (Dovidio & Fazio, 1992), where some types of thinking are more thoughtful and others more unconscious. But also, the way information is processed can vary for example between being memory-based or on-line (Hastie & Park, 1986). Judgements can draw on stored information or be constructed on-line. Increasing emphasis is placed on the context in which processing occurs (Branscombe, 1988; Linville & Carlston, 1994; Kolers & Roediger, 1984). Context can influence what information is processed, and also how it is processed. A further development is the increasing focus on more naturalistic memory with the development of research into autobiographical and everyday memory (Cohen, 1989; Rubin, 1986). Thus, in this view the focus is on explaining phenomena by reference to the mental processes and structures by which they operate (Hamilton, Devine & Ostrom, 1994). The type of information processed, how it is structured, and the type of processing engaged in is important, and all are seen as being interrelated. Increasingly attention is given to the context of processing, and to the processing of more everyday experience.

In this view what is 'in there' can take various forms, from relatively discrete instances to more generic structures that can be retrieved for judgement. But also, it is not necessarily assumed that there is anything in there; judgements may be ad hoc constructions. Getting information out, similarly, can involve a variety of types of processes, and these may interact with the current context. The question of what is 'in there' and if and how whatever is in there can be 'got out' is largely an empirical question.

Attempts to understand response effects within this theoretical framework are related more generally to the way people respond to (survey) questions, ie the psychological processes of responding. Response effects become not just problems in a survey, but psychological phenomena in their own right. The goal is to understand the response process. A first step in this direction was taken by speculating about the general processes required in the production of a response to a question. The process of responding to survey questions has been broken down into several subtasks (Tourangeau 1984, 1987; Cannell 1985b). Four general processes have been proposed: comprehension, retrieval, judgement and response.

First, Tourangeau and Rasinski (1988) suggest, respondents need to understand what the question refers to; what it means. Thus the meaning of a question is not a given but is rather a process engaged in by the respondent. Secondly, respondents must recall or reconstruct relevant information from memory. The information accessed to answer a question may rely upon a

variety of memory processes. Thirdly, the question may call for a judgement requiring a manipulation or combination of retrieved information. Judgemental processes such as information integration or a variety of heuristics may be used. Finally, the respondent must select a response. In part, this may involve mapping an answer onto a response scale. A second aspect of this fourth stage is editing the response, for example, in terms of self-serving biases and consistency with previous responses. While these stages in the response proces are described sequentially, it is assumed that there may be feedback loops from later to earlier stages.

I now want to look in detail at research aimed at understanding survey response from the psychological perspective. To do this I divide the area by the type of question being investigated, namely questions that ask for reports of behaviour and those that ask for reports of attitudes, opinions, or other subjective states. I will start with a review of work on behavioral questions. This review is meant to give a general flavour for the lines of research herein. I will then review work on attitudes, opinions, and other subjective states. This starts with a review of the nature of attitudes and continues to look at work in CASM (Cognitive Aspects of Survey Methodology) exploring the reporting of attitudes. I then want to draw out some general points about the research in the CASM programme, both with respect to behavioral and attitudinal questions and responses.

1.3 BEHAVIORAL QUESTIONS

Surveys often include questions that ask respondents to provide estimates of some aspect of their behaviour. Variously termed behavioral frequency questions or quantitative autobiographical questions, these type of questions are aimed at measuring how often respondents engage in a behaviour. They range from reports of infrequent behaviours (eg. number of dental visits, or number of crime victimisations) to relatively mundane behaviours (eg. amount of alcohol consumed or hours of television viewed), and, from more objective to subjective behaviours (eg number of household appliances purchased or number of times pain is experienced).

There are also differences in the kind of report requested. Some questions call for very accurate estimates of the frequency of a behaviour in a given period; others may ask for the rate of occurrence (eg once a month) rather than absolute frequency. Others might ask for more general estimates of the rate of occurrence using vague quantifiers (eg sometimes, rarely). Some questions ask about the respondent's *usual* behaviour, and some questions ask for comparative estimates of behaviour, for example, behaviour now versus in the past, or compared to other people.

It is assumed that, at least in principle, there can be a correct answer to these questions; people have actually engaged in a certain amount of behaviour, which could,

theoretically, be accurately counted. However, usually there are no other corroborating sources of information; the person's report is the only source of information. Thus the accuracy of the report is more central for these types of questions.

1.3.1 Strategies for producing behavioral frequency estimates

The application of cognitive theory to measuring behaviour in the survey begins by looking at memory. Theories of memory play an important role in examining behavioral questions. Expectations about the types of strategies used to respond, how accurate the answer is, and where sources of inaccuracy are likely to come from, depend a great deal on the model of memory used. Ideally if a respondent is asked to provide a response to a question about how often a behaviour occurs the respondent should recall and count each incidence of the behaviour over the relevant time span. This is known as episode enumeration. This implies a search and retrieve model of memory. Memories must be assumed to record an event and store it as it happened. Retrieval involves having the appropriate cues to access the record, and often these cues are not available, hence people can fail to retrieve potentially available material, so called errors of ommission.

This model of memory does not however, appear to be an entirely appropriate one. Tulving (1983) distinguished between episodic and semantic memory. The latter involves general knowledge. The former is the type of memory assumed to store events or episodes in detail. Thus, it would be this

type of memory that would be associated with recall of behavioural episodes. However, episodes tend not to remain at the level of discrete events. Neisser (1986) describes how event memories are nested within molar memories. Invariant properties of episodes are abstracted across repeated events and these abstractions become more accessible than the individual events from which they were derived. Others see this shift from instance to generality as a transition from episodic to semantic memory (Linton, 1986). Both, however, agree that with repetition discrete instances become difficult to recall.

However, it is not simply the lack of retrieval cues and the indistinguishability of different episodes which can cause problems. Memory involves not only retrieval but reconstruction (Bartlett, 1932; Mead, 1934). Memories are not simply stored but can be constructed from schemas and from the contexts in which one encodes and retrieves the event. In this model memories can change in both content and form; it is possible to 'recall' events which never happened. The task for those investigating survey response is to examine how such constructive or schematic memory mechanisms influence reports of behaviours. And, furthermore, it leads one to look at other types of processes which may be occurring. If respondents are not simply retrieving episodes, they may be using other types of strategies to infer the amount of a behaviour. These strategies need to be identified and evaluations of the accuracy of these strategies determined.

Much of the work done on behavioural questions has examined estimates of the absolute frequency of a behaviour within a particular time span. These types of questions are often of particular concern to survey researchers. The main types of errors which occur with these questions are omissions of relevant behaviours, underreporting, and the incorrect inclusion of irrelevant behaviours, overreporting, or errors of commission. Omissions are most commonly seen as due to retrieval failure whereas incorrect inclusions are often due to the misdating of an event that occurred outside of the given reference period. This is usually due to forward telescoping where events appear to have happened more recently than they actually did.

For episode enumeration the respondent needs the time to recall and appropriate cues for recall. Given time and repeated attempts respondents can retrieve more instances of the desired behaviour from memory (Means, Mingay, Nigam, & Zarrow, 1988). This poses a problem for survey research where there is often little time to contemplate an answer. It is also likely that providing appropriate recall cues can increase the number of events recalled (Strube, 1987). The problem here is finding appropriate recall cues. The best cues are those which match encoding circumstances, yet little is known about how everyday events are encoded. Social and personal factors may be important in determining appropriate recall cues. Particular strategies used for searching memory have also been found to affect the quality of recall. Searching from most to least recent event was found to be

better than the reverse (Loftus & Fathi, 1985). Presumably this is because more recent events provide cues for the recall of more distant events. However it was found that the latter, sub-optimal strategy was more commonly used.

Dating events is also problematic. Survey researchers often bound the recall period with calendar type dates. Yet often this type of information is not linked with the memory for an event (Strube, 1987). Bounding the period with public or personal landmark events has been shown to improve recall (Loftus & Marburger, 1983).

However, it is not simply the case that some episodes will be left out, but that other strategies will be used to estimate the amount of behaviour. For example, Smith, Jobe and Mingay (1991) reporting research on dietary memory, suggest that respondents often report on the basis of generic knowledge, what they typically eat, rather than on the basis of specific memories of what was actually eaten. This type of reporting seemed to increase for longer retention intervals. Means and Loftus (1991) looked at the recall strategies used to answer questions about medical and dental visits which varied in their likely frequency. They found that recall of individual events decreased with the number of events reported whilst rule based estimation strategies were more likely to be used for more frequent events and, to some extent, for dental visits which may be a more regularly occurring event.

Thus estimations of the amount of a behaviour rely not only on the recall of episodes of that behaviour but also rely on inference strategies to generate estimates. Use of generic knowledge, for example, relies on inferences from normal behaviour. Thus as well as looking at memory and how the structure of memory may lead to errors in reports of behaviour, one also needs to look at the types of inference strategies people use and how these may lead to bias.

Based on theories and research in other areas of cognitive psychology, a number of possible strategies have been identified. These include strategies based on a number of judgemental heuristics as well as more theory based strategies.

Judgemental heuristics

There are a variety of types of judgmental heuristics which may be used for answering behavioural questions. A decomposition strategy breaks the problem down into different elements. For example, if asked to report the number of shopping trips made, respondents might break it down into type of shopping trip, food shopping, clothes shopping etc. This type of strategy is likely to lead to inaccuracies when the categories used for decomposition are not exhaustive. However, it might provide a method for improving behavioural reports. Means and Loftus (1991) found estimates of health visits improved when respondents were encouraged to use time-line and decomposition techniques.

With the availability heuristic the ease with which an instance comes to mind is taken as indicative of its frequency of occurrence (Tversky & Kahneman, 1973). Problems associated with availability stem from the overrespresentation of salient instances.

With anchoring and adjustment, the respondent picks a number (perhaps based on the response alternatives or on some other judgemental strategy) and 'adjusts' due to recognised deviations. The problem here is that adjustments are usually not enough; the anchor tends to over-influence the estimate (Tversky & Kahneman, 1974).

Bradburn, Rips and Shevell (1987) report that decomposition strategies have been observed in responses to survey questions, as well as anchoring and adjustment. However, they do not report conditions or question types which influence the use of these strategies, nor how prevalent these strategies are. Burton and Blair (1991) compare the use of episode enumeration and other strategies. They found that those reporting fewer target behaviours used more episode enumeration; a longer reference period produced less enumeration, and giving more time to respond produced somewhat more enumeration. They report the use of strategies which involve a combination of episode enumeration and adding an amount to that recalled on the basis of the availability heuristic. Strategies based on rates of occurrence and direct estimation were also observed. Their results are also suggestive of a differential use of episode enumeration

depending on the type of behaviour being measured, eg characteristics such as distinctiveness and regularity.

These results echo those of Smith et al (1991) and Means and Loftus (1991). The more behaviours there are to count the less likely they will be enumerated and the more likely some other judgemental strategy will be used. This may reflect both a lack of cognitive effort by respondents, and the schematic nature of memory organisation.

Theory based inference

Other types of inference processes rely more on people's theories of the world. In estimating past and present behaviour (and attitudes) people may rely on their present position and the use of subjective theories of stability and change (Conway & Ross, 1984). Their estimates will be based on whether they believe the behaviour, or the self in regard to the behaviour, is likely to be stable or to have changed.

General world theories or normative expectations may also affect reports. Pearson, Ross and Dawes (1992) discuss research on stressful events. They suggest that most past research which has relied on retrospective reports of stress over long time periods found differences between men and women because people's reports were influenced by their theories of how men and women deal with stress and what they find stressful. A study by Hamilton and Faggot (1988) which asked about stresses the previous day found no such sex related differences, providing some support for Pearson et al's

proposition. That is recall of recent stressful events, a simpler memory task, showed less use of schematic models than recall of more distant stressful events.

Lay theories of memory may also play a part in estimates of behaviour. If an instance is not recalled a respondent may decide that the failure to recall is simply due to the difficulty of recalling a particular instance rather than being due to the instance not having occured.

There is a body of research dealing with the strategies people use to report behavioural frequencies and under what circumstances they use them. In order to investigate which strategies are actually used it is necessary to use some kind of process tracing technique (such as think-aloud) and there are few studies which apply these techniques to survey responses, in part because of uncertainty about verbal protocols (Burton and Blair, 1991). The studies reported above are unusual in this regard.

So far then we have seen that the time allowed for response and the type of behaviour questioned can influence the type of strategy adopted. Another aspect of the response situation which can influence behavioural reports is the questionnaire itself. Aspects of the question which can influence responses are its wording and the response scale or categories used.

1.3.2 Response scales

Schwarz and colleagues have concentrated on the informational value of response scales in responding to behavioral frequency questions. Through a number of studies they have demonstrated two effects of response scales (eg Schwarz, 1990; Schwarz and Bienias, 1990).

Firstly, response scales may help respondents interpret the meaning of a question. This is most likely to be the case when questions address a vague, or subjective, behaviour. For example, respondents were presented with the question 'how often are you annoyed' with either a high or a low frequency scale (Schwarz, Stack, Muller, & Chassein, 1988). Respondents who received the high frequency scale reported more incidence of annoyance than those respondents who received the low frequency scale. A 'meaning shift' is hypothesised to occur because respondents given the high frequency scale use the scale to infer that the behaviour in question is trivial, and therefore common, whereas those given the low frequency scale interpret the behaviour as being more severe and, therefore, less common. In line with this, when asked to provide examples of a typical example of an annoying experience, those given the low frequency scale reported more severe annoying experiences than those given the high frequency scale.

O'Muircheartaigh, Gaskell, and Wright (1992) have extended this work to include more substantive topic areas, and less extreme scales, using large scale surveys. Initially using high and medium frequency scales, with questions on annoyance

with television adverts, feeling unsafe where they lived, and experiencing physical pain, they failed to produce the large effects which Schwarz et al (1988) had produced. A further experiment varied priming of the topic area (primed or not primed), as well as frequency range. Priming questions increased the effect of scale frequency. Thus it appears that, although differences in frequency scale affect response, the effect is magnified by priming the area.

A second way in which response scales may have an impact on responses, is by informing respondents about the range of a behaviour in the population (Schwarz, Hippler, Deutsch & Strack, 1985; Schwarz & Bienias, 1990). With unambiguous questions, respondents may infer that the response scale represents the range of a behaviour in the population. This may, or may not affect their response to that particular question, but it may also effect responses to later questions. For example, Schwarz et al (1985) have shown that giving either a high or a low frequency scale for 'amount of television viewed' affected respondents' subsequent reports on questions pertaining to satisfaction with leisure time, and how important television was to them. In particular, those who reported on a low frequency scale, thus making them seem to view more television than average, reported television as being more important and reported less satisfaction with their leisure time. This 'comparison shift' appears to result from respondents taking the scale to represent a sort of usual distribution. Upon responding they place themselves within

- 41 -

that distribution and inferences about other behaviours follow from their position in that distribution.

Thus, response scales may both inform the respondent as to the meaning of the question and provide information about how the respondent compares to others. Response scales should be viewed as part of the question, effecting in some cases the interpretation of the question to which they are attached, and in some cases, affecting responses to subsequent questions, in some cases both effects may occur.

This research illustrates how respondents construct responses on the basis of the information provided in the situation. People are not simply retrieving episodes but are rather constructing an estimate.

1.3.3 Summary of behavioural research.

Research on behavioural questions has shown that retrieval of behavioural episodes is often a difficult task for respondents. In many cases they do not retrieve either episodes or estimates of behaviour, but rather construct estimates on the basis of a variety of inference strategies. Features of the questionnaire, such as the response scales, may provide information to respondents that influences their judgements. In this way aspects of the questionnaire influence reports of behaviour.

The type of behaviour in focus is also important in determining the strategy used. Behaviours can be frequent or

infrequent, regular or irregular (other dimensions may also be relevant). Different processes may be used to elicit estimates of different types of behaviour. What 'type' a behaviour is, is likely to depend upon social and personal habits. That a behaviour is mundane or unusual is in part a social fact. Because our culture has, in general, ready access to televisions, television viewing becomes a mundane activity. If we lacked such access this behaviour would be an infrequent, more memorable occurrence. At this level then the characteristics of behaviour are socially defined. But within this, there will be individual variations in what type a behaviour is. For people in our culture who lack access to a television, television viewing may still be an infrequent and more memorable occurrence. The importance of this for survey measurement is that the type of behaviour we are interested in is not manipulable, but rather we need to be sensitive to the social characteristics of the behaviour, and possible group and individual variations in the behaviour.

1.4 ATTITUDE QUESTIONS

Attitude questions are also very important to surveys. These types of questions differ from behavioural questions in that they are more subjective, and it is uncertain whether there is any accurate answer to these questions. In this section I will briefly review theories of attitudes before going on to review research within the CASM framework which focuses on context effects on attitude questions.

1.4.1 The nature of attitudes.

The conception of what an attitude is will affect how it is measured, and the interpretation of that measurement. In this section I want to look at conceptions of attitude in terms of their implications for measurement. In particular I want to look at what, according to the different conceptions, produces the response to an attitude question.

Early attitude theorists saw the attitude as a kind of general guiding framework, or mind set, that might underlie many activities such as learning, perception and judgement. Allport (1935) traces this usage back, in part, to the notion of einstellung, or set. This and related concepts (eg aufgabe) represent an overall mental preparedness or positioning in a task. More specifically, in social psychology, especially through the work of Thomas and Zaniecki (1918), the attitude was conceived of as a state of mind of the individual towards some stimuli or social object. The attitude predisposed a person to evaluate objects in a particular way.

To measure an attitude social psychologists needed to get at this general state of mind. To do this they would sample from the 'attitude universe' and select items tapping various aspects of the attitude. The attitude was inferred from the covariation amongst responses to these items (the more behaviouristic theorists regarded the attitude simply as the consistency among responses). The point was to sample enough of the relevant attitude universe to be able to say something

about the underlying attitude, rather than merely saying something about the responses to particular items. Responses to attitude items were assumed to be produced by the attitude. Whilst this model did not necessarily rule out the possibility of temporary or dynamic attitudes, in practice attitudes tended to be regarded as relatively enduring and stable -- as dispositions.

There were attempts to define more clearly the structure of attitudes. Rosenberg and Hovland (1960) proposed that attitudes were composed of affective, cognitive and behavioral components. Thus in measuring an attitude researchers needed to sample items from the different components. They also suggested that the structure of an attitude may affect its stability. In particular Rosenberg (1960) found that low affective-cognitive consistency was a feature of attitude instability.

One problem of the tripartite structure model is that it assumes a relationship between the three components. A one-component view of the attitude was proposed by Fishbein & Ajzen (1975). In this view attitudes are seen as the evaluative component involved in behaviour. Cognitive and behavioral components are taken out of the attitude concept; beliefs are seen as influencing attitudes which in turn influence behavioral intention (along with other factors). But this model also has been questioned as presupposing a relation between attitude and cognition and behaviour.

Behaviour for example has been seen as an antecedent of attitudes (Bem, 1967; Kelman, 1979).

More recently, the focus has been on attitudes as representations in memory involving associative networks. In this later view there is often less emphasis on the structure of attitudes per se, and more emphasis on general cognitive structure. Whilst the attitude still retains its essential nature as an evaluation, it is sometimes treated simply as an evaluation based on various elements in the cognitive system, and sometimes these elements are seen as part of the attitude.

Zanna and Rempel (1988) treat attitudes as evaluations. Attitudes involve 'the categorisation of a stimulus object along an evaluative dimension'. This categorisation is based on cognitive and affective information as well as information about past behaviour and behavioral intentions.

Pratkanis and Greenwald (1989), regard attitudes as being represented in memory by an object label, an evaluative summary, and a knowledge structure supporting the evaluation.

Tourangeau and Rasinski (1988) also regard attitudes as structures residing in long term memory, however there may be no pre-existing evaluation to draw on but rather the attitude / evaluation may be constructed from elements retrieved at a given time. They are not specific as to whether these elements are an essential part of the attitude or whether the

attitude is to be regarded as the evaluation based on these elements.

Whatever the disagreement about the exact structure of attitudes in memory, however, these various conceptions generally agree that the stability of attitudes is in question. They regard attitude stability as an empirical question. The conception of the cognitive system is a dynamic one. Expression of an attitude involves associations amongst a variety of cognitive elements. What attitude is expressed at any given time depends on the particular associations made. Attitudes are often temporary constructions (Wilson & Hodges, 1992). Even when an evaluation can be retrieved, exactly what is retrieved and how it is expressed is likely to be influenced by the measurement context (Feldman & Lynch, 1988). Thus, attitude measurement is no longer seen as simply tapping what is already there, but as a process of interaction between cognition and the measurement situation; a process that may involve constructing an attitude 'on-line'.

All these recent attempts to redefine attitudes recognise the importance of context in looking at attitudes. Most however conceptualise the relationship between attitude and context by treating the two as separable entities -- there is an attitude and there is the context in which it is reported, or constructed. Palmerino, Langer and McGillis (1984) take the relation between context and attitudes further. Drawing on Lewin's field theory (1939), they regard attitudes as a relationship between the person and an object. Thus the

attitude is not seen as residing in the person or being held by the person, but rather is defined in terms of the relationship. They define an attitude as 'a relation *between* two entities [person and object]. Furthermore, these two entities not only are attached to one another but the person and the object become part of an extended structure that is the context.'

The conception of context as outside the individual reflects the largely individualistic conception of attitudes. In earlier conceptions the attitude simply lies in the individual to be retrieved when required. Later conceptions view the individual as in a more dynamic relationship with his or her environment, and yet the social nature of that environment is largely neglected. Often social aspects of attitudes are acknowledged and are often implicitly used. Many researchers draw on issues that have a particular social structure or base their research on content drawn from media sources. Hovland, Harvey and Sherif (1957) for example in a study on communications and attitude change specifically use an issue which is controversial in a given area, use social groups which have a particular stand on the issue and draw different sides of the communication, in part, from media sources. All these factors draw upon the social structure and content of the attitudinal issue -- an issue can only be controversial at a social level. Yet rarely is the social structure of attitudes and its relation to individual attitudes explicitly and systematically explored.

Thus whilst recent conceptions may neglect some aspects of attitudes, the view of attitudes proposed is different from earlier conceptions. The idea that there are 'real' attitudes, or that attitudes are stable dispositions is undermined. Attitudes may or may not be drawing on a prior evaluation. Measuring attitudes involves the interplay of cognitive contents, structures, and processes, within a measurement context. Attitudes may be constructed within this context based on whatever considerations are, temporarily, accessible (Feldman & Lynch, 1988; Tourangeau & Rasinski, 1988; Wilson & Hodges, 1992; Zaller & Feldman, 1992). In this view then, the measurement context is important for understanding attitudes.

1.4.2   Context effects and attitude surveys.

Whilst some research on attitude questions in the CASM framework has examined wording effects (Hippler & Schwarz, 1986; Rasinski, 1989) and effects of response scales (Schwarz, Knauper, Hippler, Noelle-Neumann, & Clark, 1991; Schwarz & Wyer, 1985; Krosnick & Alwin, 1987), most research concerning attitude questions has focused on the context in which the questions are asked. My focus will be on context effects.

Assimilation and Contrast effects

In terms of the direction of the context effect there is a distinction between assimilation and contrast effects. That is effects in which the target is answered in line with the context and effects in which the response to the target goes against the context. This is merely an observation of the way

in which context effects are manifest. If the context has positive implications for the target then we would expect the target to be more positively evaluated; if the context has negative implications for the target then we would expect the target to be more negatively evaluated. Both of these are assimilation effects. If the opposite occurs we have a contrast effect. Whether an effect is assimilation or contrast depends on the a priori implications of the context for the target.

It is unclear why context sometimes produces assimilation and sometimes produces contrast. Attempts to explain assimilation and contrast effects in surveys draw on theories from other areas of social cognition. Research in these areas sometimes has a slightly different focus.

Sherif and Hovland (1961) for example, used the terms assimilation and contrast effects to refer to two processes. On the one hand these effects were defined in terms of whether a statement was assimilated towards the person's own stance on an issue or whether it was contrasted to it. The person's initial position was used as an anchor in judging attitude relevant material. On the other hand they also used the terms to refer to positive and negative attitude change. Assimilation, or positive attitude change, was change in the direction proposed by the communicator. Contrast, or negative attitude change, was change in the direction opposite to that proposed by the communicator. The terms as used in CASM are closer to this latter definition.

Essentially, there are two ways in which assimilation and contrast effects are investigated. On the one hand, stemming from psychophysics, the research model used is to have subjects rate different stimuli along the same dimension. For example, when judging the weight of objects, it is found that when an extreme anchor is used judgments of later weights are contrasted towards this anchor. If a heavy weight is given first a medium weight will be judged as lighter than if a light weight had been presented first (Brown, 1953). Similar results are found with less psychophysical stimuli, for example with judgements of the typicality of beer as German drink (Schwarz, Munkel, & Hippler, 1990) and with the importance of issues (Sherman, Ahlm, Berman, & Lynn, 1978).

On the other hand, there are studies that tend not to look at ratings on the same dimension but to prime different dimensions, or categories, and look at the ratings of targets when different dimensions, or categories, are primed. For example, Higgins, Rholes, and Jones (1977) primed positive or negative trait items (eg. adventurous vs reckless) as part of an 'unrelated' task. They then had subjects rate how favourable they were to a target person described ambiguously with regard to the primed dimensions. People primed with positive trait items rated the person more favourably than those primed with negative trait dimensions. That is, they assimilated to the context.

Whether these different research approaches are dealing with distinct phenomena, involving different processes is unclear.

They tend to be dealt with together at a theoretical level. Herr, Sherman, and Fazio (1983) note that usually assimilation effects due to priming are found (as in the impression formation work), whereas, in much of the social judgement work contrast effects are found.

Martin (1986; Martin & Achee, 1992), looking at assimilation and contrast effects found in impression formation studies, proposes a set/reset model to explain these effects. Use of accessible information is likely to result in assimilation. However, processing objectives determine whether accessible information is used. Information is not used if it is seen as inapplicable to the judgement. In terms of previous context, if subjects are aware of a previous, irrelevant, priming episode they are likely not to use this information in forming a judgement. When information is suppressed because it is not relevant to the current processing objectives, the subject 'resets', partialling out the primed information and searching for other information on which to base a judgement, often resulting in contrast. In priming studies, contrast has been shown when subjects are aware of the priming episode, whereas assimilation is shown when they are not aware of the priming episode. However, contrast occurs when the prime is not relevant for the judgement (eg simply stereotypical traits, as opposed to individuating information). When the prime is relevant for the judgement it may lead to assimilation, even though people are aware of it.

Lombardi, Higgins, and Bargh (1987) propose a similar model of assimilation and contrast effects. In their view though, awareness has the effect of making the primed concepts seem more extreme. Thus contrast results not from a partialing out of information, but from comparison with an extreme standard.

Schwarz and Bless (1992) theorise that assimilation and contrast effects occur as a result of categorisation processes. Simply put, for assimilation effects to occur a later item must be included in the same category as the context items. For contrast effects to occur the context must be excluded from the representation of the later item. When contrast effects are found one of two types of processes may have occurred. First, the context may simply have been excluded or 'subtracted' from the target item. Second, it may have been excluded but then used as a standard of comparison against which to judge the target item. Thus inclusion is necessary for assimilation; exclusion is necessary for contrast.

Tourangeau and Rasinski (1988) relate assimilation and contrast effects to the various stages in the response process. Effects at the comprehension stage essentially involve the resolution of ambiguity. Assimilation may occur when the respondent infers that an ambiguous question relates to the previous context and uses this context to interpret the meaning of the later question. They see contrast effects at this stage as being of the specific - general type. Respondents interpret a general question on an issue asked

after a specific question on the same issue to mean 'excluding the specific area just mentioned'. They use the context to infer the range of what is meant to be included. The target issue itself is not ambiguous, indeed Tourangeau and Rasinski suggest that the effect is most common for familiar issues. Rather, in this case it seems the context itself causes ambiguity about what should and should not be included.

The key to response effects at the retrieval stage is accessibility. Material retrieved for answering a previous question is accessible, and is therefore, due to cognitive economy, likely to be used in answering a later question. This results in assimilation when prior context makes considerations favouring one side of a related issue accessible. Contrast effects are more problematic. Accessible material may not be used, but if it is not used it may simply not lead to an effect. When the exclusion of accessible material pushes the respondent in the opposite direction contrast occurs. Why this leads to contrast is not explained cogently. The assumption is that respondents reject or discount the accessible material because they have become conscious that it is invalid or irrelevant to the target question. This is assumed to be a more controlled process than the processes producing assimilation.

In terms of judgement, prior context can serve as a standard of comparison. Assimilation can occur when the previous context suggests the dimension on which to make the evaluation. Contrast effects can occur when the context is

extreme. Self- presentational factors are seen as being most important to effects at the response stage. A desire for consistency can lead to assimilation effects. A desire to appear moderate, not to seem extreme or unreasonable, may lead to contrast effects.

Some problems with the models. While these models go some way towards explaining assimilation and contrast effects, there are still some aspects which remain unclear.

First, there are problems with a simple inclusion/exclusion view of assimilation /contrast. One experiment by Schwarz and Bless, which purports to show how exclusion leads to contrast actually shows an assimilation effect, possibly involving exclusion. Specifically, an experiment was conducted concerning ratings of the German Christian Democratic Party. The figure of Richard von Weizacker was used as context. Richard von Weizacker is, apparently, a well respected politician and is both a member of the Christian Democrat party and president, an apolitical position something like the Queen. Prior to rating the Christian Democrats, respondents in one condition were asked to say which party von Weizacker was a member of; in another condition subjects were asked to recall his apolitical position first, and in another condition there was no contextual information about von Weizacker. Ratings of the Christian democrats went up relative to the no context condition when von Weizacker's party identification was previously made salient, and went down in relation to no context when his apolitical position was made salient. The

first condition is identified as an assimilation effect due to inclusion of a positive context example in the overall category and the second as a contrast effect due to exclusion of a positive example. However, both effects are in the direction expected by the researchers. That is, both effects are in line with the implications of the context and, by this definition, are assimilation effects. A contrast effect would have occurred if the reverse results would have been found -- that is if highlighting party membership decreased ratings of the Christian Democrats and highlighting apoliticalness increased them.

If this is not a contrast effect, but rather an assimilation effect, then we might assume that the exclusion of context information from a category can result in assimilation. Thus a simple separation of assimilation and contrast into inclusion / exclusion based on categorisation is not possible.

I believe the confusion of interpretation in the above experiment is due to a confusion of the two strands of research on assimilation and contrast mentioned above -- social judgement and priming. That is rating items along a single dimension versus priming related areas. In fact it may be better to see these instances as dealing with different types of effects, perhaps involving different processes.

Secondly the idea that simple discounting or partialing out of context is necessary for contrast is problematic. Whilst it may be necessary, it does not seem to be sufficient.

Partialling out of contextual information could simply lead to no effect at all. There must be some way in which the partialled out information leads to the opposite implications. Schwarz and Bless attempt to deal with this by suggesting that the subtracted information must be more extremely valanced than the information left. This may be part of the answer, but perhaps not a complete answer. For example, if people are aware of a positive prime why not simply discard this and use a different positive piece of information? Why opt for, opposite, negative information? Perhaps there is a sense in which the discarded information implies or evokes its opposite. Perhaps contrast effects are more likely where opposite information is more readily evoked. This would seem to be most likely when associations between negative and positive information are strong.

Factors influencing context effects

As well as explaining the direction of the effect which context can produce, there is also a need to understand why the effects occur at all. The focus of research on context effects for attitude surveys has been the search for factors which make these effects more or less likely. A number of factors have been found, or are hypothesised to produce context effects, either assimilation or contrast effects. Most have focused on priming related areas. For the purposes of discussion, these can loosely be classified as question factors, situation factors, and person factors.

Question factors

Factors within the questionnaire which have been shown or which research suggest may produce context effects include: positional factors, relatedness of context and target, and type of question in terms of issue familiarity.

Positional effects. The position of the context in relation to the target may influence context effects. This is assumed to work through accessibility. Prior context is more accessible if it has been primed more recently. The further away the priming episode, the less impact it will have on the target.

Some studies have shown reduction in context effects due to buffering -- placing non-related items between the context and the target. Other studies, however, have not shown such effects (Bishop, 1987).

Tourangeau, Rasinski, Bradburn and D'Andrade (1989a, 1989b) looking at the effect of related contexts on a variety of opinion questions, found that placing context questions in a block immediately preceding the target produced larger context effects than when the context is scattered amongst unrelated questions prior to the target. However, context effects were still present when the context was scattered. The context effects Tourangeau et al found were in the direction of assimilation. However, one question produced assimilation effects when blocked and contrast effects when scattered.

However, it may also be the case that placing the context immediately prior to the target is more likely to lead to contrast effects because the context is more blatantly related to the target and hence respondents are more likely to become aware of this relationship. Evidence for this type of blatant priming effect comes from other areas such as impression formation. Lombardi, Higgins, and Bargh (1987) for example found that when subjects were made aware of a priming task they showed contrast in their judgements, but assimilation when they were not made aware.

Ottati, Riggle, Wyer, Schwarz, and Kuklinski (1989) looked at placement of specific items in relation to general items. They found that when the specific item was separated from the general item by a number of questions the general item was assimilated to the specific item whereas when it was immediately preceding the general item a contrast effect was found.

Relatedness. For context to effect the target, it must not only be accessible but must also be perceived to be related to the target. Context can be both conceptually and episodically related to the target. Here I am concerned with conceptual relatedness. Feldman and Lynch (1988) discuss the 'diagnosticity' of previous questions for the present question. This is the extent to which the respondent perceives some implication of previous questions for the present question. The more diagnostic previous items are,

they suggest, the more likely this context will produce an effect.

There is certainly evidence from priming studies to show that context must be conceptually related to the target for context effects to occur. For example Higgins Rholes and Jones (1977) primed positive, negative or unrelated traits in a previous, unrelated task. Later, when forming an impression, those given related traits assimilated the target to the traits, those given unrelated traits showed no effect. Thus accessible information was used, but only when applicable.

Tourangeau, Rasinski, Bradburn, and D'Andrade (1989b) use the correlation between context and target as a measure of relatedness, presuming that items which are highly correlated are closely connected in memory. They found greater context effects when context and target items were highly correlated.

These studies then suggest that assimilation is likely when context and target are more closely related. Is the conceptual relatedness of issues likely to produce contrast? There does not appear to be any studies directly addressing this question, however, there may be a link. Recall that contrast is said to be likely when people become aware of a priming episode (Martin, 1986; Strack, Schwarz, Bless, Kubler, & Wanke, 1993). People are most likely to become aware of the priming episode when it is blatant. But what makes priming blatant? In most studies it is, seemingly, the episodic relation between prime and target, but perhaps conceptual

relatedness may also play a part in making people aware of the priming context. That is a prime which is conceptually related in a very direct, or blatant, way to the target may be more likely than a more subtle conceptual relationship to induce an awareness of the priming episode, and thus may be more likely to lead to contrast.

Issue familiarity. Tourangeau and Rasinski (1988) suggest that unfamiliar issues may be interpreted in terms of the previous context, thus leading to assimilation effects. There is some evidence of assimilation effects when people are presented with unfamiliar or fictitious issues (Bishop, Tuchfarber, & Oldendick, 1986; Schuman & Presser, 1981; Strack, 1992).

Tourangeau, Rasinski, Bradburn, and D'Andrade (1989) report that the context effects they found did not depend on issue familiarity. However, they present no measure of issue familiarity, using policy issues likely to be present in the media, which, while some may be more familiar than others, may not vary substantially in their familiarity.

Situation factors.

I am grouping under this heading some rather different types of factors. Factors within the situation that have been shown, or suggested to produce context effects include the time given to think and the motivation to think -- the effort people put into thinking, and the conversational context of the questionnaire.

Thinking effort. At a general level many theories in cognitive and social psychology consider people to be 'cognitive misers' (Taylor, 1981), or that much social behaviour is mindless (Langer, 1978). Krosnick (1991) has described how people may satisfice in answering attitude questions. Rather than engaging in optimal strategies for answering questions, people will satisfice and produce an acceptable answer by the simplest means rather than an optimal answer which may involve more complex processing.

However, the amount of cognitive effort devoted to responding is influenced by a number of factors. In general these may be considered to include the actual time available for thinking and the motivation to think.

The first factor here is straightforward. Where there is little time to think, less thinking and less complex strategies are likely to be used. Where there is more time, there is more likelihood of more thinking and the use of more complex strategies. Surveys are generally characterised by time pressure and are therefore more likely to result in low cognitive effort.

Motivation to think may involve a number of factors, generally including the importance of thinking and interest in the topic. Factors such as the perceived consequences of 'getting it wrong', the goals of the person, and affective investment may all play a role in the perceived importance of engaging in thinking, but so too will interest in the area.

Cognitive effort is often assumed to be related to assimilation and contrast effects. Little effort is assumed to be more likely to produce assimilation effects whilst more effort is assumed to be required for contrast effects (Martin & Harlow, 1992). In many circumstances less thinking seems to promote a greater effect of context. This is because prior items, being salient, offer a simple, and therefore quicker answer to the question. This would lead to assimilation; the respondent selects the first thing in mind.

It has been suggested that a contrast effect can be produced when people become aware of, and explicitly reject the previous context, this is thought to require a greater degree of thinking (Martin, 1986). Some evidence for this is provided by Martin, Seta, & Crelia (1990). In an experiment on impression formation they blatantly primed certain concepts and then asked people to form an impression of the person described ambiguously with regard to these concepts. Some subjects, however, were played a distracting tape while forming their impressions, which should reduce the amount of cognitive capacity they have for forming impressions. Subjects who were not played the distracting tape showed contrast, whilst those who were played the tape showed assimilation.

Martin and Harlow (1992) discuss a number of factors which may influence the amount of cognitive effort exerted in response to survey questions. In looking at the effect of difficult or easy filter questions on reports of political interest, they

show that failure to answer questions leads to brooding and an attempt to distract oneself from the topic of failure. These subjects seemed to expent less effort on political questions and more on unrelated questions than did subjects who were successful and therefore basked in their success by thinking about the topic of success.

They also suggest that because informed respondents may find it more enjoyable to answer questions on which they are informed they may exert more cognitive effort than uninformed respondents. Furthermore, any differences in response may be due to differences in effort rather than differences in attitudes. They also suggest that respondents may exert more cognitive effort in face-to-face interviews than in telephone interviews, at the beginning rather than at the end of long interviews, and when they see their individual responses as important rather than simply being one of thousands.

Conversational norms. In many ways conversational norms are extra-situational. They are not specific to the questionnaire situation but are general conversational rules. By comparing the questioning situation to a conversation, research has shown that norms guiding conversational interaction may help explain context effects (Strack, Martin & Schwarz, 1988). Other areas of psychology have also made use of conversational rules in explaining behaviour; for example, research in attribution theory has shown that psychological processes, in this case explanations or attributions, may be influenced by

conversational norms (Hilton, 1990; Slugoski, Lalljee, Lamb & Ginsburg, 1993).

One particular conversational maxim that has been studied is the idea that people should make their contributions informative, avoiding redundancy (Grice,1975). Research has shown that if asked a specific question about marital happiness before a general question about overall happiness people report being less generally happy. However this seems to occur only when the two questions are seen as being part of the same conversation; the question about general happiness is taken to mean 'apart from marital happiness' (Strack, Martin, & Schwarz, 1988; Schwarz, Strack & Mai, 1991).

Martin and Achee (1992) describe this as another way in which people partial out irrelevant information in forming judgements. That is, use of the information is seen as inappropriate to current processing objectives.

Norms of cooperation in conversation may also influence response where a question is ambiguous. Here, as in ordinary conversation, people use the context of the conversation to interpret ambiguous information (Strack & Martin, 1987; Schwarz & Strack, 1988).

Person factors.

Many individual difference factors might be associated with context effects (eg sex, educational level, age etc). However, I will not be discussing demographic distinctions. Here, I

want to discuss individual differences on attitude dimensions. Among the person factors examined as relevant to context effects are: attitude 'strength' or structure, and prior knowledge.

Attitude strength/structure.   Whilst the structure of attitudes is uncertain, a number of dimensions, included in the concept of attitude strength, may be relevant to their measurement, and may also interact with context. Scott (1968) identified a number of attitude properties which could be regarded as dimensions of attitude strength: magnitude, intensity, ambivalence, salience, affective salience, cognitive complexity, overtness, embeddedness and flexibility.

In a paper arguing for the routine measurement of strength dimensions in surveys Krosnick and Abelson (1992) discuss five measures of strength which are easy to measure and extensively validated: extremity, intensity, certainty, importance and knowledge. Extremity refers to the degree of favourableness or unfavourableness of the evaluation of the attitude object. It is generally conceived of as a departure from the neutral point of the measurement scale, and is thus perhaps the most common dimension on which attitudes are measured. Intensity is the strength of the affective response to an attitude object.   Certainty is the extent to which a person is confident that their attitude is correct.   Krosnick and Abelson see ambivalence as related to this dimension, however, Olson and Zanna (1993) treat this separately from attitude

strength. Importance is the degree of personal importance the person attaches to an attitude or attitude object. Knowledge is the amount of information and beliefs linked to the attitude. Generally, these dimensions are only weakly correlated with one another (Raden, 1985). Krosnick, Boninger, Chuang, Berent, & Carnot (1993) found a number of dimensions of attitude strength were only weakly correlated. In fact, their research suggests that while the term 'attitude strength' may be a useful shorthand for referring to the various dimensions of attitudes, it does not appear that these dimensions form one underlying dimension of attitudes.

I will deal with dimensions of knowledge and direct experience in a separate section below, as these are often dealt with separately from other dimensions of attitude strength. Now I will look at other dimensions of attitude strength and how they may be related to response effects.

Krosnick and Abelson discuss a number of ways in which 'strong' attitudes may differ from weak ones. For example, attitudes which are more extreme or more important have been shown to be more resistant to change (Osgood & Tannenbaum, 1955; Rhine & Severance, 1970). There is evidence that more extreme attitudes, important attitudes and attitudes about which people are more certain are more consistent with behaviour (Fazio & Zanna, 1978; Krosnick, 1988; Sample & Warland, 1973). Thus dimensions of attitude strength seem to be an important aspect of attitudes.

Dimensions of attitude strength have also been implicated in retrieval processes. Krosnick (1989) measured latencies along with importance and concluded that important attitudes have faster response latencies. His reasoning was that important attitudes are more accessible and hence answered more quickly. Roese and Olson (1994) however, show that accessibility may be used by respondents to infer how important an attitude is in line with self-perception theory (Bem, 1967). Whichever way the causality goes, and it may be different in different cases, it is clear that accessibility and importance are positively related.

Precisely how attitude strength might influence response effects is, as yet, unclear. It was assumed that people with strong, or crystallized attitudes would be less susceptible to response effects than those with weak attitudes, sometimes seen as non-attitudes (Payne, 1951; Converse, 1974). However, evidence for this has been equivocal.

Krosnick and Schuman (1988) report on a series of studies which show context effects for a variety of question factors including question order, wording changes, and response order. Measures of attitude 'strength' were either of attitude intensity, certainty or importance. They found as many results supporting the idea that 'strong' attitudes are less susceptible to response effects as results disconfirming this idea. The only reliable effects were found for comparisons between questions which offered a middle response alternative and those that did not. People with less intense attitudes or

unimportant attitudes are more likely to opt for a middle alternative if it is offered. Similar results were found by Bishop (1990).

Krosnick and Schuman suggest that the mechanisms underlying response effects which they tested -- that response effects result from changes in perceived extremity of response options, changes in self perception, or changes in attitude -- may not be the mechanisms underlying response effects. Thus measures of attitude strength may still be important, but may interact with different underlying processes not tapped by the types of questions they addressed. Another possibility is that the measures of strength they used -- intensity, certainty and importance -- are less clearly related to response effects than other measures. It is possible then that different dimensions of strength may be related to different processes producing response effects.

Tourangeau, Rasinski, Bradburn and D'Andrade (1989a, 1989b) investigate the possibility that context serves to prime beliefs which then influence responses to subsequent questions. They found the largest context effects for those respondents whose attitudes are both important and conflicted (i.e. ambivalent). Importance alone had little effect while ambivalence alone produced slightly more effect. They argue that this is due to the person having both the necessary links between target issues and context items and having both sides of the argument available. If they are not conflicted, but their attitudes are important then they will accept only one

side of the argument and thus context becomes less influential.

Prior knowledge. Prior knowledge, experience or involvement with an issue can influence context effects. Feldman (1992) notes that one important factor which influences whether prior items will be used to form a response to a question is the extent to which alternative inputs are available. The availability of alternative inputs depends, in part, upon the person's prior knowledge of an issue. This would, by itself, suggest that more knowledge of an issue would make people less susceptible to context effects, however, the relationship is not that straightforward.

Attitudes which are automatically activated upon an encounter with an attitude object are likely to be those formed through direct experience and are more likely to be linked to behaviour (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). In this case there is a stored evaluation strongly linked to the attitude which is spontaneously, and effortlessly, recalled upon encounter with the object. Such automatically activated attitudes guide the interaction with the attitude object. If we regard the survey question as representing the attitude object, then one would expect these attitudes also to be less susceptible to context effects, because it is the evaluation which serves to respond. One can view this as a continuum between spontaneous attitudes and non-attitudes. Some attitudes have linked evaluations, but the associations are not strong enough to be automatically accessed. Other

attitudes are not formed at all. In this case the person may deliberatively process information to arrive at an attitude. In these cases the context is likely to have more impact on the production of an attitude.

More knowledge of an issue, however, could work to make people more susceptible to response effects. Recall that for context effects to occur it is thought that the context must be seen as related to the target issue. If people with more knowledge of the issue link that issue with more related issues, then it is likely that more contexts will be seen as related to the target issue, and thus are more likely to have an influence. Questions which are less generally familiar or questions preceded by a somewhat subtle context may be more open to this kind of effect.

summary of factors

A variety of factors have thus been examined and shown to be related to context effects. The relative recency of research effort in this area means that there are still many questions left unanswered. How the various factors might interact and how much they generalise across different questions, for example, are two issues which have received relatively little attention. There is still much work to do to understand the intricacies of how these factors operate, and how generalisable they are.

1.4.3 Summary of attitudes

The more recent conceptions of attitudes which are drawn on by researchers in the CASM framework, regard attitudes as knowledge structures. For some attitudes there may be stored evaluations, others may involve the on-line construction of an attitude. Thus attitudes are sometimes temporary constructions.

Much of the research in CASM has focused on understanding context effects. Context effects differ by the direction of the effect -- assimilation or contrast. Efforts to explain why effects go in different directions have drawn on more general work in social cognition, and a number of different explanations have been offered. In general the explanations see accessibility and applicability, along with processing objectives, as the important factors in these effects. Awareness of the priming nature of context and perhaps the amount of cognitive effort is also seen as important in producing contrast effects. It was argued that there is sometimes confusion over the direction of the effect stemming from different research paradigms and that to date most of the explanations offered seem to provide only partial explanations.

Identifying the factors which contribute to context effects has been the main focus of research in CASM. A number of different types of factors have been reviewed. Here, again there is much work needed to understand more clearly how the various factors operate.

1.5 CONCLUSIONS

The assumptions on which survey measurement is based, that people can give meaningful answers to questions, is questioned by findings that responses are affected by a number of seemingly trivial alterations in questions. The result of these troubles has been an effort to apply psychological theory to survey response and a considerable amount of research has been conducted in this framework.

Psychological theory does not assume that people can give reliable answers to questions. The situation is rather more complex. Sometimes people can retrieve answers; sometimes they must construct responses. What people have 'stored' in memory and how what is stored is accessed varies due to a number of factors. In part, there is the cognitive system itself. Information is often simply not stored in a way which allows direct retrieval of the desired information. In addition a variety of factors impact on this system and on responding. Situational factors such as the specific construction of the questionnaire, and the time allowed for response influence responding. So too does the conversational context of a survey interview, which can be guided by a number of conversational rules. This latter 'situational factor' draws on social conventions in interaction.

Other social factors tend to receive less focus in the CASM model, and perhaps should be given more attention. The type of behaviour or attitude one is dealing with, as noted

earlier, is in part a function of its social structuring. Whether a behaviour is frequent or infrequent, whether an attitude issue is highly structured or has never been previously encountered, depends, in part, on the individuals location in a social structure. Mapping this social structure of the issues one is dealing with may prove to be an important element in understanding response. A few studies suggest this is the case. Tourangeau, Rasinski, and D'Andrade (1991) for example, map the clusteral structure of attitudes and show that whether a context item is drawn from the same or from a different cluster affects response time to the target. This is not one individual's structure, but an aggregate level structure. Wright, Gaskell and O'Muircheartaigh (1994) point to the importance of the level of the group when they show that the meaning of vague quantifiers varies as a function of one's group membership. These studies point to a social level of structuring impacting on individual cognition.

However, one must be careful not to lose sight of individual differences. Individual differences have been shown to be influential in the effects of context on attitude reports. That is, as well as the social structure of an issue, the individual's particular structuring of the issue is also important. Less research on behavioural questions has addressed issues of individual differences. Schwarz and Beinas' (1990) relating self-awareness (Fenigstein, Sheier, & Buss, 1975) to behavioural reports is an exception here.

All these factors combine to suggest an explanation of survey response as a highly complex process. This process involves an individual, with a particular kind of cognitive system, placed in a time constrained situation governed by interactional rules, an individual who uses cues within that situation, as well as his or her own individual and social backgrounds and current processing goals, to dynamically retrieve and construct responses. No more 'ask a simple question, get a simple answer'. Once we start examining the survey response process we discover a micro-social system with all the complexities that entails.

The CASM research has certainly illustrated the complexity of the response process. It has also begun to identify factors relevant to the explanation of response, but there is still a lot to be done. In part, some under researched factors need more emphasis, but also, even those factors which have received much attention still require further elaboration. One of the challenges to this research is the development of methods which allow the in depth exploration of response processes. In part, some of the limits of explanation stem from a lack of direct data on processing. Much research relies on split-ballot experiment. However, along with taking on the theories of cognitive psychology survey researchers are also exploring its methods. One of these methods, verbal reports of cognitive processes, looks promising but is controversial, even within psychology. The next chapter takes up this issue.

CHAPTER 2

VERBAL REPORTS

SUMMARY

This chapter reviews the status of verbal reports as data on cognitive processes. Research using verbal reports as data on the processes of answering survey questions is considered first. The discussion continues by examining verbal reports as data in psychology more generally. The historical debates over introspection are reviewed as well as recent criticisms and defences of verbal reports.

"verbal reports, elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable source of information about cognitive processes." (Ericsson and Simon, 1980, p 247).

"it would appear that people may have little ability to report accurately on their cognitive processes" (Nisbett & Wilson, 1977, pp246)

.

2.1 INTRODUCTION

The status of verbal reports as data in psychology has fluctuated from one of acceptance to rejection and most stages in between. Currently, there is much debate about the role of verbal data in psychology, as the quotes above illustrate. Attempts to use verbal reports as data about cognitive processes are often challenged about the validity of different

types of verbal reports. This is especially true when collecting verbal reports in areas where they have not been previously applied. The survey questionnaire is one such area.

Verbal reports may provide a potentially rich source of data for the understanding of survey response processes. To date, much of the research looking at cognitive processes underlying survey response has involved indirect measures of processing, normally split-ballot experiments designed to produce differences in the pattern of responses. However, there is also a need to collect more detailed and direct information on cognitive processes. Verbal reports offer this type of information, but the controversy and confusion over their use suggests a need to carefully examine this method.

In this chapter I will examine the use of verbal reports as data on cognitive processes. I start with a review of their use in survey research; this illustrates both the potential of this method, but also the uncertainty surrounding its use. Because of this uncertainty, I then want to look in more detail at the use of verbal reports in psychology; first, placing the controversy surrounding them in an historical perspective and then moving on to more recent developments in the use of verbal reports.

2.2 VERBAL REPORTS AND SURVEY RESPONSE

Attempts to gain insight into respondents' understanding of survey questions pre-dates the interest in linking cognitive psychology with survey methodology. A number of studies have collected verbal reports; often, however, many psychological assumptions underlying the collection of these reports are neglected or left unexamined.

Belson (1981), for example, developed a "re-interview" technique to investigate question understanding. The day after an initial interview, he re-interviewed respondents, asking them to think back to the way that they had answered the questions the previous day, and prompting them for their understanding of key words etc.. Clearly there are problems with regarding this data as a measure of what people were thinking while they were responding. First, the time lag between responding and reporting on responding means that there may be serious recall problems. Secondly, the fact that respondents have answered a number of questions before reporting on any specific question means that there may be interference between responses to different questions. Thirdly, the use of specific directive probes means that there are problems with the validity of the data.

Schuman (1966) used a random probe technique to look at respondents' understanding of survey questions. The interviewer probed for respondents' interpretation of a random selection of questions using non-directive probes. This

technique does not have the recall and interference problems in Belson's technique in that the probe occurs directly after the respondent has answered. This is more likely to provide a measure of what respondents were considering while responding than Belson's method. The use of non-directive probes also offers advantages over directive probes, although some of the probes actually used might elicit explanations of 'why' the respondent answered in the way s/he did.

The studies above use verbal reports to improve questionnaire design. They do not attempt, per se, to examine the cognitive processes behind responding, but rather they intend to provide information on problems with specific questions that will in turn lead to improvement in these questions. They do not attempt to answer the question "what are the cognitive processes underlying response to this question" but rather "is this a good question". Sykes & Morton-Williams (1987) suggest that two forms of verbal reports, interaction-analysis and Belson's re-interview, be routinely used for pre-testing questions. Evidence from these studies suggests that verbal reports would be very useful in this regard. However, their use in this way does not necessarily give a measure of cognitive processing during response.

With links between cognitive psychology and survey methodology, greater use has been made of verbal report techniques in survey design notably on several large US surveys (the Consumer Expenditure Survey, Lessler, 1989; the National Health Interview Survey, Means, 1989; and the Labour

Force survey, Dippo, 1989;). Willis, Royston, and Bercini (1991) report on the use of verbal reports in the cognitive laboratory. Whilst they are interested in cognitive processes, and are certainly informed by cognitive theory, the main purpose underlying their use of verbal reports is in their practical use for survey design. This can be clearly seen in the way the methods are used and the criteria suggested for the assessment of their validity.

Willis et al concentrate on a technique which they call verbal probing. This technique involves the interviewer probing, either generally or specifically, immediately after the respondent has answered. An advantage of this technique, which they point out, is that it does not interfere with ongoing processes (as might thinking aloud during response); however, the drawback is that it relies on recall. Willis et al fail to point out a particularly difficult problem with probing, especially directive probing, namely that it may itself influence reports of processing, if indeed it gives information on processing during responding at all. An example of a probe which they give is "What does the term abdomen mean to you". It is unclear here whether this will generate what the respondent had in mind as they answered a question with the term 'abdomen' in it, or whether it will generate the respondent's general knowledge of the term. In the first case, the dynamics of knowledge use is being examined, in the latter, simply knowledge. An understanding of the latter can be informative for the questionnaire

designer, but it is less useful when one is trying to understand how knowledge is actually used in responding.

The quality of the data produced is also problematic. For example, it may be important to find out how sure a respondent is of their answer, but a simple 'not very ' in response to the probe 'How sure are you of your answer' is weaker evidence than if the respondent had spontaneously said 'well I'm not really sure but..'.

Willis et al's interest in verbal reports, shared with predecessors not so informed by psychological theory, focuses on the practical implications for survey research. They suggest that the validity of verbal report methods will lie less in its theoretical merits than in whether or not it reduces survey error. I would suggest, however, that there are two problems with this position. First, valid as measures of what? As the example above shows, some reports may produce a valid measure of general knowledge, but poor measures of cognitive processing during response. Measures of general knowledge may reduce survey error (eg by eliminating poorly understood or ambiguous terms) but that does not make them valid measures of processing. Second, I would suggest that reduction of survey error is only part of the goal. In some cases, a good understanding of the cognitive processes involved may show that 'error' is endemic to some kinds of questioning; the focus of concern then becomes accurate interpretation of data.

The techniques developed by survey researchers have practicality as their central concern; the objective is to produce 'better' questionnaires. While the techniques described above may be useful for these immediate goals, they may be problematic for testing and building general theory.

Some researchers have focused more on the cognitive processes underlying response as the main purpose of their use of verbal reports. In general these researchers have employed some form of protocol analysis, as described by Ericsson and Simon (1980,1984), to surveys. These techniques involve respondents either thinking aloud while responding -- that is saying what they are thinking-- or giving an account of what they thought immediately after they have responded to a question. Sometimes a combination of concurrent and retrospective reports is collected. In these techniques there is no probing from the interviewer, and reports are given as close to the time of response as possible. These techniques will be discussed in greater detail later in this chapter.

Loftus (1984) reports the use of think aloud to generate data on the cognitive processes underlying response to factual behavioural report questions concerning visits to the doctor. She suggests that the data are useful. However, she notes that the more relaxed reporting procedure of think aloud, as opposed to a formal checklist procedure, may have given respondents more time to think about their responses and, thus, to recall instances of the behaviour in question,

suggesting a change in processing due to the think aloud procedure.

Bishop (1989) used think aloud followed by immediate retrospective reports to examine some well known wording and context effects in attitude surveys. He concluded that the think aloud technique can give useful information both for generating, and for testing, hypotheses about response effects. In the case of the latter especially, he notes that large samples may be necessary. This is because, for some respondents, processing may be so automatic as to leave no trace in the verbal protocol. However, the extent to which larger samples would alleviate this particular problem is questionable. Some processes may be so automated that they do not appear in protocols at all. Thus, the range of thinking that can be assessed with verbal protocols needs investigation. Bishop gives little indication of the validity of the information produced, but rather accepts that since it has been successfully used in some areas it can be applied to the survey. This may be a problematic assumption in regard to context effects, where if, as Loftus suggests, respondents have more time to think, they may also better remember previous questions and thus context effects may be increased.

Ursic and Helgeson (1989) used think-aloud to look at the decision making processes involved in questionnaire responding. They varied questionnaire length, scale anchors, and number of scale points in order to look at differences in processing among these questionnaire factors. Respondents had

to think-aloud while rating fast food restaurants. The data resulting from these protocols appears interesting and suggest a number of hypotheses about questionnaire decision making. Unfortunately, details of the instructions used for think aloud are not given, nor do Ursic and Helgeson appear to have investigated whether think aloud itself has any effect on response. Whilst respondents were found to be accurately verbalising their thought processes, this information appears to be obtained through debriefings in pre-tests to develop a coding frame, and therefore may be a dubious measure of the validity of the verbal reports.

Burton and Blair (1991) used retrospective protocols to look at the processes that were used when respondents answered behavioural frequency questions. The retrospective report method was used because as Burton and Blair say it "allows the frequency question to be asked and answered in a manner natural to the survey; concurrent process measures would change the task and, possibly, the process used in response formulation." (Burton & Blair, 1991, p67). Thus, they have some doubts about the validity of think-aloud.

Burton and Blair reported two studies. The first, looking at estimates of number of B grades achieved and courses taken outside the university, used written retrospective protocols. The second, looking at the frequency of withdrawals from automatic teller machines, used telephone interviewing and asked subjects for retrospective reports with the prompt "How did you come up with that answer?"; probing by interviewers

was used to clarify process descriptions. The written protocols obtained were sometimes ambiguous and not precise enough for hypothesis testing. This might be expected to be the case from previous research (eg. Byrne, 1983). The data from the verbal retrospective reports does, however, appear to be useful in testing hypotheses about the use of episode enumeration. Again, Burton and Blair gave no data concerning the extent to which these reports are valid.

An earlier study (Blair and Burton, 1987) did provide some reasons for believing that retrospective reports of processing during response may be valid. Retrospective reports of thinking were gathered for a question asking for frequency of eating at a restaurant. Different conditions were hypothesised to effect the amount of episode retrieval, specifically a longer reference period was hypothesised to produce less reference to episode enumeration than would a shorter period. Indeed, evidence supporting this was obtained in the protocols. Blair and Burton regard this difference in the protocols as evidence that the protocols were valid. This seems to be the most convincing evidence so far in the survey literature that verbal protocols can provide valid information on processing during responding.

Thus, the use of verbal reports within survey research has different goals. Some see it as a means for dealing with problems in questionnaires; largely their aim is to improve specific questionnaires. Others see it as a way of collecting data on cognitive processes; largely their aim is to

contribute to a theory of survey response. Most of these studies suggest that the data provided from verbal reports is useful, to both aims. Yet there is also a good deal of uncertainty about what kind of data the different methods provide, and about the validity of different methods. Given that this debate over the adequacy of verbal reports extends to psychology in general, it seems useful to examine, in some detail, the history and current controversies surrounding verbal reports. I start this review by looking at early uses of verbal reports in psychology, namely introspection.

2.3 INTROSPECTION AND ASSOCIATED VERBAL REPORTS

Introspection was a favoured method of early psychologists. It was seen as giving access to thought, to the mind. However, disagreements over its validity, and more importantly the move to a behaviouristic science, with its emphasis on observable behaviour and disregard for consciousness, led to a decline in the use of introspection.

Burt (1962) called for the reintroduction of introspection as a method in psychology. In criticising psychology for taking behaviour rather than consciousness as its defining characteristic he says of the then present day psychology "having first bargained away its soul and then gone out of its mind, seems now, as it faces an untimely end, to have lost all consciousness". It seems that, more recently, psychologists have regained consciousness and rediscovered mind. Since the 1960's psychologists have shown increased interest in consciousness (Valentine 1992).

Along with these theoretical concerns there have been a number of reassessments of the place of introspection as a method for psychology (e.g. Joynson, 1972; Radford, 1974; Lieberman, 1979; Howe, 1991). Increased interest in mental activity has led for a need to access, in some way, the stream of consciousness. Getting reports from people on their mental activity seems one way to do this. But, with introspection having ben discredited, how can this be done? First, I think it is necessary to look at the method of introspection, to understand what was wrong with it, before proceeding to look at more recent uses of reports of mental events which are more or less similar to introspection.

2.3.1 Early introspection

Wundt established his experimental psychology using introspection as the method. Wundt was sophisticated in his use of introspection. He, recognised the distorting nature of introspection and, thus, attempted to control introspection by rigorous experimental procedure. He states that:

> "First ... mental processes may not be observed while they are taking place. We must limit ourselves to analysing them, so far as possible, from the effects which they leave behind in our memory. Secondly: wherever it is possible, we must endeavour so to control our mental processes by means of objective stimulation of the external organs ... that the disturbing influence which the condition of observation tends to exercise upon them is counteracted. This control is given by experiment" (Wundt 1894/1977, p12-13).

Recognising the distorting nature of introspection, ie. that in observing a process the process, by this observation, is altered, Wundt attempted to overcome this problem by rigorous experimental procedure. Observation requires the focusing of

attention. Yet, for introspection, because the observer and the observed are one and the same, this is an impossible task without changing that which is observed. For Wundt, the experiment provided this focus of attention. Very simple objects were presented (such as letters, or points of light), and often only for a short time, so that, essentially what was available to be 'ideated' was limited, and, therefore, focused. It was the careful and systematic control over the stimuli (the experimental method) which made scientific introspection possible (Wundt 1904). Allied to this tight control over the stimulus, observers were also trained in introspection. Partly, this approach was expected to produce such habitual observation so as to diminish self consciousness (Lyons, 1986), but probably it was also necessary so that sensations could be correctly defined.

Even with this control over the conditions of introspection, Wundt believed that much of mental life could not be studied through introspection. Whilst Wundt believed that the contents of sensory perception could be examined through introspection, he placed the study of higher cognate phenomenon in social psychology. Blumenthal (1975) equates the former, attributes of experience, with what today would be termed 'human information-processing capacities'. The 'products of common mental life' (Wundt, 1904) could not be examined through introspection, because the mind itself was a product of these forces and, therefore, the mind could not be conscious of them (Farr, 1983). They became clear only at a collective level.

Titchener followed Wundt in using introspection as the prime method of psychology. However, Titchener extended the range of phenomena to which introspection could be applied. And, whereas Wundt concentrated on laboratory conditions as the controls on introspection, Titchener laid more emphasis on the subject as the control (Lyons, 1986). Hence, he created a number of rules about what could and could not be reported in order to guard against the 'stimulus error'. That is, in introspecting, one must make no reference to the external object, all reports must be of subjective sensations. Koffka (1924) gives an example of comparing weights where the subject cannot say ' This weight is heavier than that' but rather must say 'My sensation of tension is now heavier than before'. He further points out that this distinction makes it difficult to explain the objective nature of perception. It is curious, in some ways, that after taking so much trouble to isolate the objective conditions these are then expunged from the report of experience.

Boring (1953) labels the introspection of Wundt and Titchener as 'classical introspection', others have labelled them 'structuralists' in line with their more general theoretical approach, (Valentine, 1978) and to differentiate them from the Wurtzburg school.

2.3.2 Introspection and the Wurtzburgers

The main difference between classical introspection and the Wurtzburg school seems to lie in to what introspection can be applied. The Wurtzburgers extended the use of introspection

to thought in general. However, there were also differences in the methods of introspection. Kulpe describes the differences between classical introspection and that practised by the Wurtzburg school thus:

> "Previously it was the rule not to obtain reports about all experiences that occurred during an experiment as soon as it was concluded, but only to obtain occasional reports from subjects about exceptional or abnormal occurrences. Only at the conclusion of a whole series was a general report requested about the main facts that were still remembered. .... However, as soon as persons trained in self-observation were allowed to make complete and unprejudiced reports about their experiences of an experiment immediately after its completion, the necessity for an extension of the previous concepts and definitions became obvious. (1964/1922, p209-210).

This passage implies a development in introspective methodologies. The Wurtzburg school seems to have been less restrictive in what could be reported during introspection. In their studies the Wurtzburgers 'discovered' the imageless thought. This was thought without imagery or feeling and therefore contradicted the structuralist doctrine. The ensuing debate over the existence of imageless thought is often taken as one of the reasons for the demise of introspection -- as introspection is one person's subjective experience, there was no way in which to determine which is the accurate interpretation when there are two conflicting reports. Did thought have to have accompanying sensations, or was thought possible without sensations?

Mandler and Mandler (1964) however, claim that the introspective protocols produced by the two groups were in fact very similar. It was the theoretical interpretation of

essentially the same data that led to such heated conflict. Wundt (1894/1977) had also, in discussing the clarity of an idea, noted that, while introspecting, as well as those ideas that were clearly apprehended, there may be a number of others which are less distinct or entirely indistinct. This type of 'half-formed' idea is very similar to the description, given by the Wurtzburgers, of imageless thought. Although certainly, at the time, the lack of agreement between different researchers contributed to the demise of introspection.

The Gestalt psychologists objected to the analytic introspection (which may be roughly equated with classical introspection, but provides a more descriptive term) not so much on the grounds that it could not provide valid data, but rather because of its relevance. Kohler (1930) objected to analytic introspection because the adoption of the psychological attitude which it entailed (as exemplified by the example given by Koffka above) changed ordinary experience into something quite different. He did not doubt that given the conditions of isolating the stimulus in such an extreme way one could decompose experience into the kinds of elements described by the analytic introspectionists. But rather, he thought that their results were so divorced from the way a person normally perceives that they were not generalisable beyond the psychological laboratory.

2.3.3 The behaviourist challenge

Perhaps the main challenge to the use of introspection came from alternative frameworks for psychology. In America this took the form of behaviourism. The Behaviourists' dethronement of introspection came not because they demonstrated the invalidity of introspection, but because they ignored it and set up an alternative frame for psychology in which consciousness, and the mind, played little or no part and therefore introspection became irrelevant.

Whilst some behaviourists sometimes claimed that mind did not exist (eg Watson, 1930/1967), the major objection to the use of introspection was that one can never really be sure about what goes on in another person's mind. Internal data is private and subjective, external data is public and objective. It was also thought that the experiment did not have the distorting influences that introspection did, however, more recently the reactivity of the experiment has been demonstrated (Orne, 1962). Questions have also been raised about the extent to which the external environment can be viewed 'objectively' rather than being socially constructed (eg. Gergen, 1982). Thus, more recently, the distinction between public and private data has become more problematic.

2.4 VERBAL REPORTS -- THE NEW INTROSPECTION ?

Behaviourists were not, however, able to banish introspection completely. Boring (1953) notes that introspection could be said to be dead, which he believed was true of classical

introspection, or it could be said to be alive and flourishing under various aliases, one of which is verbal reports.

Whilst behaviorism pushed introspection from centre stage in regard to academic psychology, it maintained a role in many somewhat peripheral areas of psychology. So, the Gestalt psychologists collected phenomenal reports, similar to the Wurtzburg method of introspection but using naïve, rather than trained, subjects who reported whatever they thought. A number of studies also used "think aloud" techniques, collecting the verbal reports simultaneously with a task rather than after completion. Psychophysics also found introspection a valuable, and indeed vital, method. Though, here, it is difficult to see in quite what way the use of introspection in psychophysics differs from much of the introspection practised by Wundt. Psychoanalysts used verbal reports (both in regard to retrospective accounts of dreams and in free association) to access the unconscious. Finally, with the gradual emergence of social psychology, the collection of post-experimental reports became common. Indeed, Orne (1962) suggested this method should be used as a means of assessing experimental artifacts.

Ericsson and Simon (1980) claim that dissatisfaction with classical introspection generalised to any use of verbal reports. So much so, they argue, that verbal reports have been used only as an heuristic rather than on a par with other types of hard data. One of the problems with this conception of verbal reports is that it discourages effort to distinguish

between types of verbal report and when and how they can be used. De Groot (1966), who used verbal protocols in problem solving, summed up this position when he said "many psychologists have a bad conscience about using introspective data; as a result, they often use them poorly".

2.4.1 Verbal reports -- Nisbett and Wilson's challenge

Perhaps paradoxically, social psychologists have provided one of the strongest recent attacks on the use of verbal reports. Nisbett and Wilson (1977) were concerned at the acceptance of verbal reports without any investigation of their validity. They attempted, in part, to evaluate more systematically the use of verbal reports; one aim was to look at why people are sometimes accurate in giving reports of higher cognitive processes and sometimes not.

Nisbett and Wilson review work in the area of attribution theory, subliminal perception, problem solving, and helping behaviour in the presence of others. It is perhaps surprising that the review of problem solving focuses on retrospective reports of 'insight' and not the numerous studies using some form of process tracing technique such as "think aloud". In all the cases reviewed they conclude that subjects have been shown to be unaware of the processes producing their behaviour. In other words, the psychologists and the subjects' explanations for the subjects' behaviour are different. As the psychologists' explanations derive from the experimental methodology, the subjects' are assumed to be

unable to accurately introspect the causes of their own behaviour.

Nisbett and Wilson then go on to conduct studies of their own, choosing situations where they felt that subjects would be unable to accurately identify the effects of a stimulus on their responses. Here again, they conclude that subjects were virtually never accurate in their reports. Their main conclusion is that, rather than drawing upon a record of the processes that have occurred, people are drawing upon lay theories of what would be plausible processes. People are accurate to the extent that their theories are accurate and not because they have introspective awareness of their cognitive processes.

This paper (and others produced in this series, eg. Nisbett and Bellows, 1977) has been criticised on a number of grounds, both methodological and theoretical (eg. White, 1980, 1985; Smith and Miller, 1978). It does not seem to be a fair test to claim that one is interested in when subjects are accurate and then conduct tests and review literature where one thinks they are likely to be inaccurate. Others have reviewed studies which they felt showed that subjects reports were accurate (eg. Lieberman, 1979; Ericsson and Simon, 1984). And, as Farr (1987) points out, Nisbett and Wilson's conclusion is not dissimilar from the position Wundt had taken in splitting his experimental, and introspectionist, psychology from his social psychology.

However, there still was no detailed analysis of why these reports are unacceptable when others seem to find introspection a useful method. Ericsson and Simon (1980, 1984), provided this much needed systemisation of the area of verbal reports. They present a theory, in information processing terms, which moves the debate from one of can we or can't we rely on verbal reports to one of under what conditions can or can't we rely on verbal reports. Because this is such an important theory it is worth describing in some detail.

2.4.2   Protocol Analysis - A brief outline of Ericsson and Simon's theory of verbal reports.

Ericsson and Simon (1980, 1984) present a theory of verbalisation in regard to internal processes. This theory aims to tackle the misuse, and lack of use, of verbal reports as data by specifying when verbal reports are likely to reflect underlying processes and when they are not. Byrne (1983) points out that verbal reports are used in two ways: 1) the content of verbal reports is analysed in order to explain other behaviour, and 2) to examine patterns and sequences in behaviour. Similarly Bainbridge (1985) says that verbal reports collected during the performance of a task give information about how knowledge is used in a particular task, whilst an interview gives information on the content, interrelations and range of knowledge. Ericsson and Simon are mostly concerned with using verbal reports for process tracing, i.e. for investigating the type and sequence of mental operations carried out by a subject performing a set

task.   Whilst their theory has implications for other types
of verbal reports, such as responses in an interview (indeed
they discuss how the theory applies to survey questions), the
focus of the theory is on tracing ongoing cognitive processes.

Their theory is based on an information processing approach to
cognition, with the basic assumption that verbal
behaviour is just an example of behaviour in general and the
cognitive processes that generate verbal reports are simply a
set of the processes that produce any other behaviour.   The
main theoretical assumptions of this approach are that
cognition is information processing, and that information is
stored in several memories having different capacities and
accessing characteristics.   Essentially, they conclude that
two types of information are accurately reportable.   Firstly,
information that is present in short term memory (STM).
Secondly, information in long term memory (LTM) for which
markers still exist in STM - these markers mediate the
retrieval process.

Based on this theory, they predict that the following types of
verbalization will be accurate reflections of cognitive
processes: 1) concurrent verbalization - that is where the
subject 'thinks aloud' during completion of a task, when
information is still in STM. 2) retrospective verbalization -
when it is conducted immediately after the completion of a
task, when information or cues for retrieving that information
from LTM are still likely to be in STM.   In fact this leads to
three possible ways to collect verbal protocols: 1) concurrent

verbalisation 2) immediate retrospection and 3) concurrent verbalisation with immediate retrospection (ie. 1 and 2 combined).

Importance of instructions to verbalise

As well as the timing of the verbalisation, other factors involved in eliciting the verbalisation also affect its accuracy. Primary among these is the instructions given to the subjects for verbalisation. Instructions are important because they can affect what type, or level, of information is reported. Ericsson and Simon describe three levels of verbalisation. Instructions to verbalise may be asking subjects to verbalise information at different levels.

Level 1 is simply the verbalisation of 'inner speech'. The information being verbalised is already in an oral code. Instructions such as: 'Try to think aloud. I guess you often do so when you are alone working on a problem' (Duncker, 1926, reported by Ericsson and Simon, 1984.) or 'Say out loud everything that you say to yourself silently' are presumed to ask subjects only to report their 'inner speech'. This kind of instruction will only be useful in tasks where one assumes that people naturally use oral codes. Where they do not, very little information is likely to be obtained from the verbal protocol.

For verbalisation at level two it is assumed that some translation into a verbal code is necessary. Verbalisation does not bring new information into awareness but involves

labelling information that is held in some non-verbal internal format. Ericsson and Simon argue that this level does not involve any change in processing, but may require more time. Instructions such as: 'Tell me everything that passes through your head' require the subject to label and encode the contents of STM and, thus, requires the kind of recoding involved in level 2 verbalisation. Note that this type of instruction may also include the reporting of level 1 -inner speech- information.

Level three verbalizations require extra processing steps in order to comply with the verbalisation instructions. In these cases the verbal protocols will not be an accurate reflection of cognitive processes because some mediating process is involved between the original thought or process and its verbal production. The subject may have to explain his/her thoughts. For example the instruction: 'In order to follow your thoughts we ask you to think aloud, explaining each step as thoroughly as you can' requires the subject not only to report but also to explain their thoughts and, thus, changes the course of ordinary processing. Subjects may also be required to report instances of particular types of contents or processes; this requires them to monitor their thought processes and thus may also change thought processes. Probes to report specific types of information would be examples of this type of instruction.

Criticisms and defence of verbal reports.
The main criticisms levelled at the use of concurrent

verbalization are: 1) It affects, and in some way changes, the cognitive processes 2) Concurrent verbalizations are mere epiphenomena, unrelated to non-verbal behaviour and 3) The verbal reports are incomplete.

Ericsson and Simon review a large amount of empirical research, mostly in the area of problem solving, where they conclude that concurrent verbalisation, "think aloud", is not likely to change processing, but may slow it down, when the requested verbalisation is at level 1 or 2 described above. However, level 3 verbalisation, as noted above, is likely to alter processing requirements. Note that this criticism is very similar to that identified by early introspectionists - namely that introspection changes the process it attempts to observe. They attempted to deal with this problem by experimental control and trained subjects. Ericsson and Simon are claiming, in fact, that there is no need for such controls. It is, rather, the type of information requested which can affect processing.

In answer to the second criticism, Ericsson and Simon argue that concurrent reports are likely to represent cognitive processes when they are at level 1 or 2. At level 3, different processes could be producing the verbal reports and thus they may be unrelated to the cognitive processes of interest. For example, if asked 'how' they answered a question subjects may theorise how they think they must have answered it. Thus the subject's theorising, a separate process, produces the verbal report. In this case verbal

reports are unrelated to the process which actually produced the original response. Where level 1 or level 2 information is provided, Ericsson and Simon give a number of ways to determine whether verbal data reflects processing.

For completeness of verbal reports Ericsson and Simon argue that 'the information that is heeded during performance of a task, is the information that is reportable; and the information that is reported is information that is heeded'. Information which is not stored in STM or not retrievable from LTM will not be reported. So such things as the process of recognition or processes which are automatic, will leave only their input and output in STM. These inputs and outputs may be reported, but not the processes that generated the output. In addition to these systematic omissions in reporting, Byrne (1983) notes that random omissions are likely, due to variations in effort and lapses in attention.

The instructions given to verbalise, may affect whether these criticisms are relevant, as they effect the level and type of information produced. So for example, to ensure that reports reflect the underlying process subjects should be asked to report only what they are thinking, perhaps as they are thinking it. They should be asked NOT to explain their thoughts. And, they should not be asked to provide particular types of information. To make reports as complete as possible they should be asked to report all their thoughts.

The collection of retrospective reports presents some of the same problems as concurrent reports. For example, asking subjects to explain their thoughts will lead to reports which are not reports of their actual cognitive processing. There are also additional problems involved in the collection of retrospective reports. Some of the main worries include the fact that subjects, if asked to report after a series of trials, may not be reporting specific information from any one trial but may be reporting generalised information abstracted from performance across some or all of the trials. Also it should be made clear to subjects that only information which they can remember attending to on a particular trial (or when answering a particular question) should be reported, that is, not to report their current thinking about how they solve a problem (or would answer a question), but what they can remember thinking when they actually solved a specific problem.

2.5 VERBAL REPORTS AND INTROSPECTION

Ericsson and Simon have detailed the conditions under which verbal reports can be said to be accurate, or inaccurate, descriptions of cognitive processes. They have provided a theory describing where these data come from, how they are produced. In so doing, they provide an explanation of what these data mean, how to interpret them. It is through this theory-based explanation that we can determine what these data are reports of and thus determine what they 'measure'.

Comparing Ericsson and Simon's theory of the production of verbal reports to the early use of introspection, it can be seen that there are a number of differences between the two methods.

At a general level, both Ericsson and Simon's 'verbal reports' and 'classical introspection' seem to be very similar -- they are both meant to be reports of immediate experience. The differences lie not so much in what they studied as in how they studied it. I think Wundt and others were, in this endeavor, not so much wrong in principle as wrong in practice. A number of differences obtain between classical introspection and the verbal protocol technique, suggesting the presence of biases in the experimental arrangements of the classical introspectionists.

Ericsson and Crutcher (1991) define the introspective approach as one which relies on reports from skilled observers; these reports are given privileged status. Protocol analysis on the other hand uses naïve respondents, and verbal reports are treated as just another form of data. Having skilled observers creates a number of problems. The subjects are not naïve to the experimental hypotheses; they may be directed as to what elements to report; and there is some confusion between data production and data interpretation, that is, these subjects not only reported their thoughts, but also analysed them. These types of reports would constitute level 3 verbalisations in Ericsson and Simon's terms, and thus pose problems for the validity of the data as reports of thoughts.

Also, classical introspection often involved reports after the completion of perhaps a number of trials. This would also raise problems with the validity of reports.

Ericsson and Simon (1984) point out that the type of experience subjects in classical introspection were attempting to report was essentially that associated with sensory stores; information not normally attended to, or available in STM, and thus difficult to report on accurately. This is, perhaps, another reason why trained subjects were necessary. Verbal reports, on the other hand, as conceived by Ericsson and Simon, should only be reports of heeded information. But, what is heeded can vary with the subject, and the conditions of observation; hence Wundt's attempts to control the information available to be processed.

Thus the use of verbal reports as data on cognitive processes may be seen as in the same tradition as early experimental psychologists. The main difference is that the theory underlying the production of verbal reports specifies different conditions for their collection, differences which influence the validity of these reports.

Since introspection's falling from favour there have been numerous reassessments of its use, and many ideas of what it is. Some have held up Freudian introspection as the model to be followed (eg Bakan, 1954; Howe, 1991), others include reports on cognitive processes (eg Lieberman, 1979), and some do not define what they mean by introspection at all (eg Burt,

1962).   Introspection is perhaps best seen as a family of methods (McKeller, 1962).   Radford (1974) defines three groups of activities which have been referred to as introspection in terms of what the subject is asked to do.   These are self-observation, self-reports and thinking aloud.   In the first, a person observes his or her own mental events.   In the second, a person reports his or her experience but without attempting to be objective.   The third is self explanatory.

The classical introspectionists may come under the category of self-observation, Ericsson and Simon's reports would be included in think aloud.   Nisbett and Ross, however, are more likely dealing with a type of self-report.   Subjects in their experiments are not engaged in self-observation, because they generally did not know, whilst engaged in the activity, that they would be queried about their thinking on it, nor did they think aloud.   Thus Nisbett and Wilson's are dealing with a different kind of verbal report than either Ericcson and Simon or the early introspectionists.

Nisbett and Wilson are asking subjects to provide reasons for their behaviour.   In Ericcson and Simon's terms they are asking for level 3 information; they are not simply reporting their processing, but explaining it.   The ability to provide reasons for one's actions rests on the ability to have insight into oneself, self knowledge.   Nisbett & Wilson seem to be suggesting that subjects should have knowledge about why they do the things they do; in fact, to such an extent, that they should be able to explain their behaviour from the perspective

of the experimental psychologists even though they do not have access to the same vital information as does the psychologist. If people did have the kind of self knowledge that Nisbett and Wilson regard as introspection, the psychologist would be redundant. It is more likely, as Nisbett and Wilson demonstrate, that people do not have this information. Rather they construct knowledge of themselves from theories and representations which they have derived through social interaction. When explaining themselves it is upon these theories that they draw.

In fact this view is very similar to a definition of introspection put forward by Lyons (1986, 1991). Reviewing early psychological and philosophical theories of introspection, he concludes that essentially these theories have given a two-level account of introspection. That is, they regard introspection as involving some second level process which monitors the first level process of thought. Thus, introspection becomes a distinct psychological process. One of the attendant ideas associated with this view is that it gives people some kind of privileged access to their thoughts. They can, as it were, observe them from above. This is the view of intropsection which Nisbett and Wilson discredit.

Lyons argues that this view of introspection is incorrect and suggests an alternative idea of what people are doing when they are said to be introspecting. In this view, introspection is simply a first order process not unlike other

first order perceptual processes. It involves replaying, reseeing, refeeling etc.. There is no need for some kind of monitoring system in this approach. In these terms introspection is a kind of self-reflection, an ability to contemplate ourselves through recollection, and imagination, and is thus part of normal psychological processing, not some kind of special activity. He sees this as being linked to self-explanation, or self-knowledge, which ultimately derives from folk theories, much as Nisbett and Wilson regard explanations of behaviour as relying on personal theories.

There is certainly some reason to doubt these types of reports as accurate descriptions of cognitive processes. Yet there may be other uses for this type of data. For example, if one is interested in lay theories of thought this kind of data would be very useful. Whether they are accurate accounts is not the issue in this context, although it may also be interesting to determine when, and for what kinds of issues, will people's accounts and psychologist's accounts converge. Perhaps the dissemination of psychological theories may alter the explanations, as Moscovici (1976) has shown to be the case with psychoanalysis. Both Lyons, and Nisbett and Wilson, argue for a theory based production of this type of verbal report. In this way they attempt to define the processes involved in the production of this type of verbal report, a definition which describes the data in terms of normal psychological processing. Thus, this kind of report, or introspection, should not be rejected as data per se, but it

needs to be treated as a different kind of data than verbal reports of immediate experience - process data.

These distinctions are important for survey methods because the survey method is also a form of verbal report, mostly involving self-reports, but sometimes respondents are asked for reasons -- Why do you do x? or Why do you think x? Nisbett and Wilson's general arguments about people's ability to explain the reasons for their behaviour suggest caution in the interpretation of such data, indeed they suggest a different interpretation of these data may be necessary.

The idea that people can give accurate self-reports is questionned in recent approaches to the survey and the processes involved in the production of these types of reports are now being investigated. In much the same way as Ericsson and Simon have described the production of verbal reports of cognitive processes, researchers in CASM are attempting to explain the production of another type of verbal report, the survey response. The need to explain the processing behind the production of this type of report makes the use of Ericsson and Simon's type of verbal report an attractive methodological option.

Looking back to the studies discussed earlier in this chapter which use verbal reports of cognitive processing during survey response, it seems that most have avoided blatantly asking for explanations of response, and thus exposing themselves to Nisbett and Wilson's criticisms of verbal reports. A question

such as 'why did you answer the question in that way' might be seen as asking for such self-explanations. Whilst most survey researchers want to obtain reports of the immediate experiences which occur while answering questions, some seem definitely to have used methods that, according to Ericcson and Simon's analysis, seriously undermine the validity of these reports as reports of heeded thoughts. Clearly, the re-interview technique poses far too great a demand on the recall abilities of respondents, and directive verbal probing is likely to involve level three processing, that is, it holds the risk of being produced by processes different from those used to respond to questions.

Some survey researchers, however, have complied with Ericcson and Simon's prescriptions for accurate verbal reporting, collecting reports during or immediately after response, without prompting from the interviewer. Having reviewed Ericcson and Simon's theory, there seems to be no definite theoretical reasons to doubt the validity of think-aloud techniques for survey response.

The empirical evidence for the usefulness and validity of verbal protocols is also generally positive. Verbal reports of cognitive processes, or verbal protocols, have been applied to many areas, aside from the area of problem solving, with apparent success. These include: decision making (Payne, Braunstein, & Carroll, 1978; Svenson, 1989; Wofford & Goodwin, 1990), applied areas such as processing engaged in by machine operators, and management appraisal (Bainbridge, 1985; Martin

& Klimoski, 1990), consumer research (Smead, Wilcox, & Wilkes, 1981; Wright and Rip, 1980), and persuasion research (Wright, 1980).

However, whilst much of the research that has examined the validity of protocol analysis has failed to show differences between verbalising and non-verbalising groups in their choices, judgements, or problems solutions (eg. Ericsson & Simon, 1984; Smead et al, 1981), some studies have found differences in processing between verbalising groups and non-verbalising groups. That is, using designs which have some subjects verbalising on all trials and some subjects verbalising only on some trials, some differences in processing has been found, for example in studies of consumer choice decisions (Biehal & Chakravarti, 1989), and impression formation tasks (Mumma, Draguns, & Seibel, 1993). These tend to suggest that concurrent verbalisation may, for some types of information, or some types of processes, facilitate task processing. This facilitation may be due to stronger memory traces due to thinking aloud. It is unclear whether these results are important in the use of think aloud in the survey. The type of tasks used in the above studies are relatively complex, much more so than the task of answering a typical survey question. It is likely that if task facilitation effects do occur, they may be greater for more complex tasks, such as choosing between a variety of items based on a number of dimensions. Survey questions are, generally, not tasks which involve the same type of complexity, though they may involve different types of difficulties (eg recall of personal

information). However, the suggestion that checks be carried out on the validity of concurrent verbalisation for different tasks (Mumma et al, 1993) seems a prudent one.

2.6 CONCLUSIONS

One of the arguments running through a number of papers calling for a reassessment of introspection is that no method is error free. This is certainly true, but it should not mean that no attempt is made to identify and deal with these errors, especially as the alternatives are either to accept error as a fact of life or to discard methods where errors are suspected. Neither of these options is conducive to the development of psychological methodology, nor by extension to the development of psychology as a whole.

In determining whether data are valid or not, it is clear that some theoretical explanation of how the data are produced is necessary. Different types of verbal reports may be produced by different processes. To determine what verbal reports are reports of one needs to distinguish what produced the report. Ericsson and Simon have provided this basic theoretical underpinning to the production of verbal reports of cognitive processes. Research within the CASM framework attempts a similar feat for verbal reports given as survey responses.

It seems that survey researchers could usefully draw on Ericcson and Simon's analysis in investigating the processes underlying survey response. Indeed, some have already done

so. But, even here doubts are expressed about their validity
and uncertainty remains over which method to use and what kind
of information they provide. Thus, an investigation of verbal
protocols in the survey seems warrented.

CHAPTER 3. EXPERIMENT 1.

AN EXAMINATION OF VERBAL REPORTS

FOR SURVEY QUESTIONS

SUMMARY

This chapter reports on an experiment designed to test
different verbal report methods for collecting data on
cognitive processes associated with responses to
survey questions. An experiment was conducted using
four different techniques for collecting verbal
reports which are in line with Ericcson and Simon's
principles for verbal reporting. An effort is made to
determine which method is most suitable for the
survey. The completeness of the reports is considered
and qualitative assessments of validity are made.

3.1 INTRODUCTION

It appears, from studies discussed in the previous chapter,
that verbal reports can be expected to provide useful
information about survey response processes. Yet, in the
survey literature there is also a good deal of uncertainty
about verbal reports: is the validity of some methods suspect?
Is information on some issues likely to be inaccessible? Even
conforming to the principles of valid reporting identified by
Ericsson and Simon, some researchers use think aloud, others
use retrospection only and some use a combination. Given this
evident confusion a study designed to assess the use of verbal

reports in surveys is warrented. This would allow a more informed choice on the most suitable method to use with surveys.

The aims of the present study are to assess which type of report -- think aloud, retrospection, or both -- is most suitable for the study of survey response processes.

This raises the question of what criteria to use to assess what is the 'most suitable' method. Ideally the method should produce verbal reports which are both valid and complete: valid in the sense that the protocol is a report of heeded thoughts rather than the method itself influencing the thoughts produced; complete in the sense that everything that is (at least consciously) thought is reported.

In terms of validity, the standard method for examining the effects of think aloud is to compare the output measures for a verbalising group and a non-verbalising group engaged in the same task. Such a comparison is not practical in the case of survey data. As there is no correct answer, large numbers would be required in order to conduct any reasonable test in this manner. The analysis of such a large number of protocols is impractical (Ericsson and Oliver, 1986, suggest that 20 is a large number of verbal protocols).

Hence, other ways of assessing the validity of the data are needed. Some suggestions of other, more qualitative, indicators of validity in verbal protocol data include: how

easy it is for respondents to verbalise (Byrne 1983), and the structure of the language used (Ericsson and Simon 1984). Comparing think aloud and retrospective reports may also provide useful information.

Given validity, the assessment of the comparable completeness of reports is fairly straightforward -- which method provides the most information. Of course this does not indicate how complete the reports are in any absolute sense, but only in comparison to other reports.

3.2 METHOD

This experiment was designed to investigate verbal report procedures for collecting information on cognitive processes underlying response to a structured questionnaire in an interview setting. To do this four groups were given different instructions for verbal reports.

Group 1 received an instruction for think aloud only, using a modified wording of the instruction suggested by Ericsson and Simon (1984) which was used by Bishop (1989). This is referred to as instruction A. Group 2 also received an instruction for think-aloud only; this was a modified wording of the instruction used by Russo, Johnson,and Stephens and reported by Svenson (1989) - Instruction B. Group 3 received an instruction for think aloud with immediate retrospection. The think aloud part of the instruction was the same as for group one, but also included instructions for immediate

retrospection. This instruction is also a modified version of that suggested by Ericsson and Simon and used by Bishop - Instruction C. Group 4 received an instruction for immediate retrospection only. This was the same (slightly modified) instruction as the immediate retrospection instruction received by group 3 - Instruction D. The full wording of these instructions is given in appendix 1. Figure 3.1 below outlines the experimental conditions.

FIGURE 3.1
Experimental Conditions

	GROUP 1	GROUP 2	GROUP 3	GROUP 4
Type of verbal report	TA	TA	TA + RETROSPECTION	RETROSPECTION ONLY
Instruction	A	B	C (A+D)	D

(TA = think aloud)

This design allows comparisons to be made between the verbal resports produced by two different think aloud instructions, between think aloud on its own and think aloud with retrospection, and between think aloud accounts, and retrospective accounts.

3.2.1 The questionnaire.

Each of the four groups received the same questionnaire. The main priorities in the design of the questionnaire were that it should be relevant to the sample, reflect the kind of questions used in surveys, and contain questions which, in theory, pose different response problems.

The questionnaire comprised nine questions. The first four questions were factual questions dealing with library use: usual weekly hours spent in library (USUAL USE), actual hours spent in library last week (LAST WEEK USE), comparing this term's use to the previous term's use (COMPARE USE), items used in library over four weeks (WHAT USE). The fifth question was a subjective behavioral question, also pertaining to the library: frequency of annoyance with library services (ANNOY LIB). The final four questions were taken from the Eurobarometer survey and were attitudinal questions dealing with issues concerning Britain: satisfaction with democracy (DEMOSAT), whether Britain has benefited from membership of the EC (ECBENEFIT), reaction to the possibility of the EC being scrapped (ECSCRAP), and what should be the priority goal for Britain (GOALS). The complete questionnaire is given in appendix 2.

3.2.2 Respondents

Forty respondents were recruited from around the LSE. Respondents were mainly students at the LSE but a few were members of staff and a few were students from other colleges who also had classes at the LSE, or came to use the library. Male and Female respondents were used. Whilst most were British, a few were foreign students who spoke very good english. The age range was from about 19 to 50. There were 10 respondents in each of the four conditions.

3.2.3 Procedure

The respondent was seated across from the interviewer. They were asked to agree to being tape recorded. None refused, or expressed any anxiety about being recorded. The respondent was then given the appropriate instruction for his/her condition and asked to read it. When he/she had finished reading the instructions, he/she was asked whether he/she understood what he/she was to do. Most said they did. Any uncertainties raised by respondents were dealt with before continuing. This consisted of the Interviewer paraphrasing the particular instruction the respondent had received, and was required only in a small number of cases.

The respondent was then given two practice questions before going on to the main questionnaire. During the practice questions respondents engaged in the appropriate verbal report procedure for their condition. Minimal feedback was given about whether they were verbalising correctly (eg "yes, that's what I'd like you to do" or "try to tell me everything you think"). After the interviewer was assured that the respondent felt comfortable with the procedure, the main questionnaire was administered.

The interviewer then asked each question and recorded answers as in a normal interview. Think aloud respondents verbalised while answering each question. Respondents in the retrospection conditions were prompted with "Now tell me all that you can remember about your thinking", at each question after respondents had answered, although occasionally respondents pre-empted this prompt by immediately reporting

their thinking. The Interviewer gave the appropriate prompt (see instructions) when long silences occurred during think aloud.

After the questionnaire was completed, each respondent was asked to report all that he/she could remember about the instructions which they had received initially. They were then asked for their reactions to and comments on the procedure.

3.3 RESULTS

These results present the analyses of the verbal protocols produced by respondents. Responses to the questions are not analysed. The data are mostly categorical and because of the low subject numbers many cell frequencies are too small for analysis.

The protocols have been analysed in a number of ways. Comparisons between the different instruction groups are made in terms of the amount of verbalisation, the similarity of content in what is said, the structure of the language used, and the amount of time taken to respond as well as the rate of verbalisation.

To compare whether the content of the protocols differs across groups a content analysis of the protocols was done. Coding frames were developed for each question using classical content analysis methods. The coding frames were developed ad

hoc from both the think-aloud and retrospective protocols. Protocols were exhastively broken down into independent statements. These statements were then given a descriptive label, for example, the statements 'I spent time on the computer [in the library]' and 'I got three books out' were identified as activities done in the library. These labels formed the basis of category development; subsequent statements given the same descriptive label were grouped together to form a category. When a sufficient number of protocols were coded to allow categories to be descriminated, further protocols were coded using these categories. Each statement was placed in only one category. Where, after coding was complete, a category had a very small number of statements, these were collapsed into more general categories or, if there was no logical relation to other categories were placed in an 'other' category. The codes produced from this analysis will also inform counts of the amount of verbalisation, and also the rate of verbalisation. (Examples of coding frames are given in figure 3.3).

3.3.1 Comparisons of amount of words and codes.

The number of words used in the protocols was compared across groups. This was expected to give some gross indication of the similarity of different instructions for verbal reports. In addition, the number of codes used (taken from the content analysis) was also compared.

Comparisons between think aloud protocols.

The mean number of words used per question in think aloud protocols is given in table 3.1 below, along with a breakdown across question type. For the behavioral and attitude questions this represents the mean of four questions; for the subjective behavioral question, the mean is, in fact, the mean for only one question. The response to the question forms part of the number of words at a question.

TABLE 3.1
Average number of words per question in think aloud protocols. (standard deviations are given in parentheses).

	Behaviour Q's 1-4	Subjective Behaviour Q5	Attitude Q's 6-9	Overall Q's 1-9
Group 1	36 (23)	70 (45)	78 (59)	57 (33)
Group 2	23 (16)	77 (48)	63 (46)	46 (25)
Group 3	35 (20)	97 (80)	110 (93)	75 (48)
Group 4	15 (12)	59 (78)	28 (27)	24 (24)

In all the groups there was a large variance in the number of words produced. Some respondents talked a great deal, while some did little more than verbalise the answer to the question. Variability seems to be due both to some respondents being better verbalisers or more keen respondents than others, but also to variation in the response task. For example, some respondents give only a response when asked how many hours they spent in the library last week (Q2) because they knew they were away and immediately say 'none'; other respondents, who did use the library, have more to think

about. Thus, the response problem is not constant across respondents.

Although group 4 respondents do not receive a think aloud instruction, some subjects do produce some verbalisation during answering the question, thus they 'spontaneously' think aloud. The variation is very large in this group.— Is it?

A one-way analysis of variance was carried out across the groups on the total number of words used for all questions across groups. This was significant ($F=4.09$, $p=.013$). A priori contrasts were carried out between the three think aloud groups and group 4; it was expected that an instruction to think aloud would increase the amount of think aloud. These results are presented in table 3.2 below.

TABLE 3.2
Results of Contrasts between Think Aloud Groups
and Group 4 (No Think Aloud)

GROUP	T Value	D.F.	T Prob.
1 VS 4	-2.628	16.3	.018
2 VS 4	-2.002	17.9	.061
3 VS 4	-3.057	13.2	.009

Groups 1 and 3 differ significantly from group 4, group 2 approaches significance. It seems that an instruction to think aloud increases the gross amount of verbalisation, however the instructions for groups 1 and 3 have a greater impact than the instructions for group 2. Comparisons

between the three think aloud groups showed that none were significantly different from each other.

Across all groups respondents tend to say less on the behavioral questions (questions 1, 2, 3, and 4) than they do on the attitude questions (6, 7, 8, and 9). Question 5 (ANNOY LIB), the subjective behaviourial question is more similar to the attitude questions in the amount of words spoken than to the behavioral questions. The fact that the attitude questions come after the factual questions suggests that the difference in amount of words between the question types might simply be a practice effect. However, the pattern across questions does not show a simple increase from earlier to later questions. (See figure 3.2 below) For example, question 8 (ECSCRAP), an attitude question, is more similar in amount of words to the factual questions. Theoretically, this is what one might expect given that it comes directly after a question on the same topic (ECBENEFIT).

FIGURE 3.2
Mean number of words per question for think aloud
groups (groups 1-3)
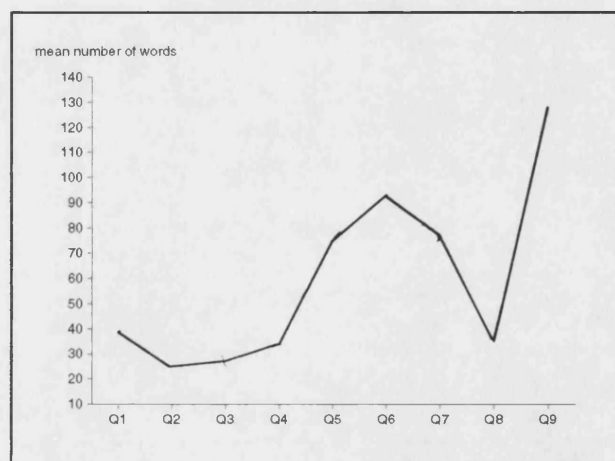
A breakdown of the average number of words per group per question is given in appendix 3. Mostly, the different groups show a similar pattern across the questions.

The number of words used provides a relatively objective measure of the gross amount of think aloud, however, it does not indicate whether what is said is meaningful. An instruction to 'talk constantly' may simply encourage respondents to use more words to say the same thing. Looking at the number of codes used, whilst less objective, gives more indication of whether the groups differ in the number of 'thoughts' produced, rather than merely the amount of words. The mean number of codes used per question type and for all questions is given in Table 3.3 below.

TABLE 3.3

Average number of codes per question in think aloud protocols. (standard deviations are given in parentheses)

	Behaviour Q's 1-4	Subjective Behaviour Q5	Attitude Q's 6-9	Overall Q's 1-9
Group 1	3.9 (2.1)	6.4 (2.8)	6.8 (2.5)	5.5 (2.0)
Group 2	2.4 (1.4)	6.9 (4.5)	5.4 (3.0)	4.2 (1.8)
Group 3	3.3 (1.4)	5.8 (3.3)	8.5 (5.7)	5.9 (2.8)
Group 4	1.8 (1.1)	4.6 (3.7)	2.6 (1.8)	2.4 (1.4)

The pattern of codes is similar to the pattern for the mean number of words. These results were analysed in the same way as above; the one-way ANOVA was significant at $F=5.79$, $p=.002$. The results of contrasts are given in tables 3.4 below. All

think aloud groups differ significantly from group 4 and none of them are significantly different from each other.

TABLE 3.4
Results of Contrasts between Think Aloud
Groups and Group 4 (No Think Aloud).

Group	T Value	D.F.	T Prob.
1 VS 4	-4.001	16.3	.001
2 VS 4	-2.485	17.0	.024
3 VS 4	-3.524	13.2	.004

A breakdown of the means for each question is given in appendix 3. The mean number of codes per question is similar to the mean number of words.

Comparison of amount of retrospection

The mean amount of retrospection per question in groups 3 and 4, broken down into question type as above, is given in table 3.5 below.

TABLE 3.5
Average number of words per question in
retrospective protocols.

	Behaviour Q's 1-4	Subjective Behaviour Q5	Attitude Q's 6-9	Overall Q's 1-9
Group 3	44 (21)	75 (63)	94 (67)	69 (37)
Group 4	63 (35)	90 (81)	84 (61)	76 (49)

There are no significant differences in the amount of words produced in retrospective reports. As with the think-aloud protocols, the variation between subjects is large. The amount of retrospection is similar to the amount of think-aloud, except that group 4 seems to say a bit more on the behavioural questions.

A breakdown of means for individual questions is given in appendix 3. Only one question produced a significantly different amount of words between the two groups; this was question 4 (WHAT USE). The mean number of words for group 3 was 39.33 (sd=29.31) and the mean for group 4 was 110.5 (sd=94.15). This was significant at t=2.27, p=.044. Group 4 respondents produce significantly more words in their retrospective reports than do group 3 respondents.

As for the think aloud protocols, the number of codes used was compared for the retrospective reports. The results resemble those for the number of words. The mean number of codes per question is presented in table 3.6 below. There is very little difference between them, and again the variance is large.

TABLE 3.6
Average number of codes per question in
retrospective protocols.

	Behaviour Q's 1-4	Subjective Behaviour Q5	Attitude Q's 6-9	Overall Q's 1-9
Group 3	3.4 (1.3)	4.0 (2.7)	6.2 (3.9)	4.9 (2.7)
Group 4	5.1 (2.4)	5.3 (3.9)	5.2 (2.5)	5.2 (2.4)

Results for individual questions are similar to those for the mean number of words. Most questions produce a similar number of codes across groups; the exception is question 4 (WHAT USE) where group 4 produces significantly more codes than group 3 (t=-2.81 p .018).

Comparison between think aloud and retrospection

The mean number of words given in think aloud and retrospective protocols was compared for group 3, where each respondent provides both a think aloud and a retrospective report for each question. The means are given in table 3.10 below.

A Pearsons correlation was carried out, on the total number of words in think aloud and in retrospective protocols, giving r= .91 which is significant at p > .001 (The correlation between number of codes gave a similar result r=.79, p=.03). Thus the gross amount of think aloud and retrospection is highly positively related, the more respondents say in their think aloud protocols, the more they say in their retrospective protocols.

TABLE 3.7
Average number of words per question in think aloud and retrospective protocols.

	Behaviour Q's 1-4	Subjective Behaviour Q5	Attitude Q's 6-9	Overall Q's 1-9
Think Aloud	35 (20)	97 (80)	110 (93)	75 (48)
Retrospection	44 (21)	75 (63)	94 (67)	69 (37)

Individual questions were compared using related t-tests, most show little difference. For question 2 (LAST WEEK USE) there is a significant difference between the amount of think aloud and the amount of retrospection; respondents say more in their retrospective protocols than in their think aloud protocols (t=-3.85 p .005). However, this seems to be due to a number of respondents who do not use the library giving only a short response at the question, but going on to say more in their retrospection.

3.3.2 Analysis of content of protocols.

Coding frames were developed for each question, because different questions asked for different information. The results presented here are thus by question, rather than across all questions. The same coding frames were used for retrospective and think aloud protocols. Examples of the coding frames are given in figure 3.3. Coding frames for all questions are given in appendix 4.

The analyses to be presented compare number of codes per category across groups, rather than the number of respondents using a code. The latter gives too little data for comparison. One problem with using number of codes per category is that it does not take account of how many respondents use the code and thus may be heavily influenced by one or two respondents' repeated use of a particular code. However, this will be taken into account where necessary when looking at the results.

Figure 3.3

Examples of Coding Frames

QUESTION 1: USUAL LIBRARY USE	QUESTION 6: SATISFACTION WITH DEMOCRACY

1. Activities
  a. In library
  b. general
 	
2. Amount of use
  a. Hours (response)
  b. number of visits
  c. Calculations of amount
 	
3. Personal statements
 	
4. About the question
  a. Repeat/rephrase question
  b. difficulties
  c. time frame consideration
 	
5. Other
 	
6. Process description
	1. Aspects of democracy
	
	2. Defining terms
	
	3. Comparisons
	
	4. Evaluations
	
	5. Dealing with the scale
	
	6. Personal statements
	
	7. About the question
	  a. Repeat/rephrase
	    question
	  b. difficulties
	
	8. Other
	
	9. Process Descriptions
	
	10. Responses

Content of think aloud protocols across groups.

Chi- squared tests were done only between the three think aloud groups (groups 1, 2 and 3), as there were too few codes for testing in the think aloud of group 4.

For most questions the distribution of codes across groups was similar. For question 4 (WHAT USE) inspection of the codes shows that group 1 and 2 respondents mention their reasons for use and amount of use of library services more than do group 3 respondents (16, 7, and 1 code used for all subjects, respectively). However, no test was conducted for this question because the low number of codes given overall would

have required the combination of very diverse categories in order to carry out a legal test. In any case, the results from this question need to be treated cautiously, as most respondents say very little, with an average of less than 3 codes per respondent. Also, the large variance means that many respondents say nothing (other than the response), hence this result is heavily biased by a few respondents who do have something to say.

Comparison of retrospective protocols.

For questions 2 (LAST WEEK USE), 3 (COMPARE USE), 4 (WHAT USE), 5 (ANNOY LIB), and 6 (DEMOSAT) no significant differences were found in the distribution of the codes. For questions 1 (USUAL USE), 7 (ECBENEFIT), 8 (ECSCRAP) and 9 (GOALS) the chi squares were significant. These are presented in tables 3.8, 3.9, 3.10, and 3.11 below.

TABLE 3.8
Analysis of content of Question 1 (USUAL USE)

	Activities	Amount of use	Other Codes	Total
GROUP 3	12	11	15	38
GROUP 4	3	30	10	43
TOTAL	15	40	25	81

X^2=15.12 df = 2, significant at .001

The differences at question 1 lie largely in the mention of "activities" and "amount of library use". Group 3 mentions more of the former and group 4 more of the latter. The former

category includes statements such as "and that I normally go in, find out on the computer or the uh slide things the books that I need...". The latter category consists of general estimates of time spent in library such as "I've been spending a lot of time in there", it also includes hours per day, days used, descriptions of calculations, and also the response of hours per week.

TABLE 3.9
Analysis of Content of Question 7 (ECBENEFIT)

	Beliefs/ Aspects	Evaluations	Process Description	personal statement	Other	TOTAL
GROUP 3	10	8	18	16	8	60
GROUP 4	9	13	8	9	17	56
TOTAL	19	21	26	25	25	116

X^2=10.25 df = 4, significant at .05

For question 7, there are a number of small differences in the use of most codes except the mention of "beliefs or aspects". The code of "process descriptions" here includes such statements as "It just sort of came in a rush".

TABLE 3.10
Analysis of Content of Question 8 (ECSCRAP)

	Beliefs & evaluations	Process Description	Other	TOTAL
GROUP 3	8	12	14	34
GROUP 4	3	3	27	33
TOTAL	11	15	41	67

X^2=11.77 df = 2, significant at .01

The differences for question 8 are mostly from the mention of more of a variety of "other" codes by group 4 and more use of "process descriptions" by group 3.

TABLE 3.11
Analysis of Content of Question 9 (GOALS)

	Mentions/ reads	evaluations	beliefs/ aspects	Process Description	Other	TOTAL
GROUP 3	12	25	14	24	13	88
GROUP 4	5	19	13	3	12	52
TOTAL	17	44	27	27	25	140

X^2=11.68 df = 4, significant at .05

The differences for question 9 are largely due to the use of more "process descriptions" by group 3, with other minor differences.

Think-aloud compared with retrospection

Comparing the codes used by respondents giving both think aloud and retrospective protocols gives some indication of the similarities between the two reports. A strict comparison of codes shows that there is a good deal of difference in the codes used. Table 3.12 below gives the proportion of codes which occur in both the think aloud and retrospective protocols.

TABLE 3.12
Proportion of codes which differ between think aloud
and retrospective protocols.

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
38	.32	.36	.17	.53	.47	.38	.37	.34

There are a number of reasons for this difference in the number of common codes. Firstly, the retrospective protocols use more process descriptions than the think aloud protocols. The 'process description' code was used only when it was purely a process description; if a unit was partially a process description, but could be coded into a different category, then it was. Two statements from question 9 (GOALS) will serve as an example:

1. "and I remember reading two and thinking that I quite liked two".
2. "..and like trying to evaluate them as quickly as possible".

This first statement was coded as an evaluation; while it partly describes thinking, it also makes a direct evaluation. The second statement was coded as a process description, because there is no actual evaluation of any response category, but only a statement that this process was carried out.

Secondly, the differences in content between think aloud and retrospection come from respondents who say very little in their think aloud protocols, sometimes only the response, but then go on in their retrospective protocols either to give reasons for the response or to describe their thought processes. Below are examples of these.

```
RESPONDENT 30   QUESTION 8 (ECSCRAP)
TA
 I'd be sorry

RETROSPECTION
 Um passports, uh because my italian passport's going through
 and I wouldn't be able to travel round freely any more
 So I'd be personally very sorry.

RESPONDENT 21   QUESTION 4 (WHAT USE)

RETROSPECTION
 Those were all very simple questions.
 and the answer immediately came to mind.
 so there were no real thought processes.
 I just knew the answer immediately.
```

Other differences stem from respondents summarising what they said in the think aloud; and some from respondents expanding on what they have said. Others, however, do seem to report different things in the retrospection than they did in the think aloud. For example:

```
Respondent 23   Question 6 (DEMOSAT)

THINK ALOUD
Okay, well I'm looking at all these numbers.  I'm looking at all these
  numbers.
Um, I think my personal satisfaction would be quite high.
Um I'm uh I think I don't know I wouldn't put ten, because that seems a
  bit extreme.  I don't know why, but it just seems a little bit much.
I think I think I'd compromise a little bit and put perhaps an eight.
I mean I'd say that was fairly high satisfaction.

RETROSPECTION
Um yeah, it was quite, just sort of quite interested, quite an unusual
  question.
First reaction was to think well no I'm not satisfied at all.
Uh but but then I suddenly remembered that I'd been travelling in
  countries where there isn't very much decent democracy.
And really although we bear a few old grudges etcetera, we do quite
  well compared to many parts of the world.
So I think you know that all things relative, you've got to give it a
  high mark because that's the way the world is.
```

For this respondent the think aloud protocol largely consists of choosing an answer, whereas the retrospective protocol reports thoughts about the question generally. It is important to note that after hearing this question the respondent has about 5 or 6 seconds of thinking time while the interviewer reads out the scale procedures.

3.3.3 Structure of language

In this section a qualitative assessment is made of the language used in protocols. The fluency of the language is assessed, as well as the kind of statements used. Also, a general assessment is made of the sequencing of statements as well as how well they relate to the question.

Think aloud protocols.

For all questions and across all groups, the language is fairly fluent and the sentences well formed. However, there are a number of repeated words, false starts and poorly structured sentences. There tends to be a similar amount of these type of sentences across all groups. Longer protocols tend to contain more of these types of sentences than shorter protocols. There appears to be individual variation in that, across questions, some subjects are consistently more fluent than are others. Below are examples of both mild and severe forms of poor sentence structure.

RESPONDENT 9 QUESTION 1 (USUAL USE)

Um depends on the days.
I'm there a couple there so it varies.
. . . .

RESPONDENT 1 QUESTION 8 (ECSCRAP)

I think I'd be sorry about it
because I like, I like, I like some of the things that Europe can give.
. . . .

RESPONDENT 22 QUESTION 7 (ECBENEFIT)

. . . .
They are, they're being, they're not being, just they want the best of both worlds.
They are comparing the costs according, I mean comparing, they're just taking the costs and not seeing the benefits.
. . . .

. . . .
It depends on whether, I well it, on whether I look things up or it depends
on whether I'm you know.
I would say it a I don't look things up every day.
. . . .

Respondents tend to address the question, ie. what is said
appears to be relevant to the question they were asked.
Protocols tend to be bound to the context, eg. there are a
number of ill specified referents which cannot be understood
without reference to the question.

Generally, the sequence of sentences seems good. Due to the
ill defined nature of the task (as noted in the introduction),
it is difficult to know what a 'good' sequence is. Responses
tend to come last, or near last, although this is certainly
not always the case. With some very short protocols, the
answer can be first with other statements following. This is
more common on factual questions where the respondent responds
fairly quickly and then backs up his/her response with a
further statement. With some attitude questions, the response
may come relatively early on but with more statements (
occasionally a lot more) after that. Repetitions of part of
the question, or rephrasing of the question, tend to occur at
the beginning of the protocol, or after a prompt from the
interviewer.

Protocols tend to consist of plain statements with very few
descriptions of processes. The latter form only a small
proportion of the total number of statements. Some of these
take the form of a sort of introduction to speaking, such as

" well my first thinking is" or "well the first thing that comes to mind".

Others describe recall processes. For example:

   "looking over last week",

   "trying to remember how much work I did last term",

   "I'm trying to think back of any special event",

Questions 6 (DEMOSAT) and 9 (GOALS) produce descriptions of response activities such as:

   "I'm just looking at the scale",

   "well reading through them".

A few describe other processes to be performed.

      "so then I have to try to figure out how much I think

            Britain is democratic".

   "I have to go to a daily and then multiply it",

   "I'm going to have to give a guesstimate" ,

A couple of statements describe processes in the past tense, but they are describing thoughts which occurred while the question was being asked, before they could think aloud.

There are only a small number of more worrying statements which seem to describe previous processes. A couple of respondents from group 3 occasionally seem to start retrospecting in their think aloud protocols. But these kinds of statements are few, and so are not a serious problem.

Retrospective reports

For group 3, retrospective reports tend to have better sentence structure than the think aloud protocols. There are fewer repetitions of words, false starts, and poorly structured sentences. The retrospection tends to consist more of process descriptions and statements phrased in the past tense. There are some present tense statements. There are some individual differences in method of reporting, in that some respondents tend to use more statements than others.

For group 4, the sentence structure is not as good as the retrospection of group 3, but not as bad as the think aloud protocols of any of the groups. It consists of a mixture of process descriptions, past tense statements and present tense statements. These latter are more common in group 4 than in group 3, making the structure of these protocols somewhere between the structure of think aloud protocols and the retrospection produced by group 3.

An example from group 3 and 4, illustrating this kind of difference are presented below:

QUESTION 2 (LAST WEEK USE)

RESPONDENT 22 -- GROUP 3
When you said last week, I just, my mind, I mean immediately went to
    the things I did last week.
Some of them were like I I was staging a play, so I thought about the
    play.
But I just distracted it because it was not related with 'library'.
I tried to separate the times I spent on my studies and my classes.
    but then it just didn't work out again.
and then at the end probably I didn't spend more than 5 minutes last
    week in the library.

RESPONDENT 33 -- GROUP 4
Well again the main thing was, as I suggested in the answer I gave, was
   just to decide whether, really whether you meant last week, what you
   meant by last week.
Sort of people generally, usually do mean different things.
They mean last actual calendar week.
Or last week as in the last week that's just gone.
Uh again since the answer would be very different for the different
   definitions of week,
it seemed sensible to clarify that for the answer.
But then after it was fairly easy.
because there was so few lectures and classes in then
That I do spend more or less the amount of time, the average amount
each day, with not too much diversion
So really that was easy.

It is worth noting that several protocols provide information
that one would not expect, mostly in group 4. For example,
one respondent changes his response on recall of another piece
of information. Another, for almost the entire retrospection,
goes back and changes his answer and describes a new piece of
information. A couple of respondents mention what they did not
think about.

3.3.4   Time taken for verbalisation.

The amount of time taken for think aloud was calculated for
each group. The measure of time was taken from the moment the
question had been asked until the moment the respondent
stopped talking. The same was done for group 4, thus their
responses and any thinking aloud they may have done was
included in the measure. The amount of time taken to read the
question was not included in the measure; at question 4 each
sub-question asked was excluded from the overall amount of
time taken to respond to the question.

Amount of time taken to respond was calculated for each
question and then summed across all questions, giving a total
amount of time spent responding for each respondent. The

means for the total amount of response time, in minutes, is given for each group in table 3.13 below.

TABLE 3.13
Mean response time - in minutes

GROUP 1	GROUP 2	GROUP 3	GROUP 4
4.35	3.41	4.79	2.68
(3.05)	(1.91)	(2.43)	(1.80)

The analysis of variance carried out was not significant. A priori contrasts showed that only group 3 differed significantly from group 4 in the amount of time taken to respond. None of the think aloud groups differed significantly from one another. The results of comparisons between the think aloud groups and group 4 are given in table 3.14 below.

TABLE 3.14
Comparisons of response time between Think Aloud
Groups and Group 4

Group	T Value	D.F.	T Prob.
1 VS 4	.803	17.9	.432
2 VS 4	.879	17.9	.391
3 VS 4	2.204	16.6	.042

The mean number of words per minute (wpm) was calculated for each group, although the rates for the think aloud groups are the most important. The mean wpm for each group is given in table 3.15 below.

TABLE 3.15
Mean Number of words per minute

	GROUP 1	GROUP 2	GROUP 3	GROUP 4
Mean WPM	125.24	116.94	134.94	68.47
SD	(25.36)	(30.97)	(16.03)	(27.47)

Whilst there is a good deal of variance between subjects in how much they say, and a fair amount of variance in the time taken to respond, the rate of speech among think aloud groups has less variance, especially in group 3. None of the think aloud groups differ significantly.

These rates of verbalisation compare favourably with amounts reported by Ericsson and Simon (1984). They report that normal relaxed continuous talking produces about 150 to 200 wpm, while rates of between 50 to 110 are reported for think aloud in problem solving studies (they do note one expert who verbalised on a problem solving task at a rate near to normal speech).

Comparisons show that all three think aloud groups verbalise at a rate significantly different from group 4. Thus, the time taken to verbalise for groups 1 and 2 is not different from group 4 but their rate of verbalisation is greater. For group 3, although respondents take longer to answer than group 4 respondents, those in group 3 are verbalising at a much quicker rate.

Respondents in the think aloud groups tend to begin vocalising as soon as the question is asked, and, as can be seen from the rates of verbalisation, are fairly fluent. Respondents in group 4 who verbalise tend not to do so immediately, but after a pause, presumably while thinking.

Words per minute is useful for comparisons with normal speech, but is less useful for comparing across tasks, where for some tasks one word represents an entire thought while for others a sentence may be needed. In the present study, the nearest approximation to a measure of 'thoughts' is represented by codes. Some codes may be only one word, while a few are made up of a number of sentences. The mean number of codes per minute for each group is given in table 3.16 below.

TABLE 3.16
Mean Codes per minute

GROUP 1	GROUP 2	GROUP 3	GROUP 4
13.51	12.31	11.47	8.63
(4.16)	(2.82)	(1.58)	(2.09)

Again, this compares favourably with amounts reported by Ericsson and Simon where in various problem solving tasks they report rates of 8 to 11.

3.4 DISCUSSION

The discussion of these results will focus on the two main aims of this study, that is 1) Which reports are most

complete and 2) How valid are the protocols as measures of ongoing processes?

3.4.1 Which report/s are most complete?

For the think aloud reports the amount of think aloud, measured both by the amount of words and by number of codes, does not vary significantly across groups. The pattern is similar for both measures: Group 3 respondents say the most, then group 1, and then group 2. There is a slight tendency for group 3 respondents to say more than group 2 respondents in both words and codes, but there is a large amount of variability in these measures.

It may be that the instruction to 'talk constantly' , included in the instructions of both groups 1 and 3, increases slightly the amount of think aloud (across both these groups this is the second best recalled part of the instruction, second only to naming think aloud). The addition of an instruction for retrospection may have the same effect, thus leading group 3 respondents to say the most.

If retrospecting does increase slightly the amount of think aloud this might be for at least two reasons: 1) Producing both types of reports may give a better understanding of what ' thinking aloud' means (Ericsson and Simon 1984) which may facilitate the production of think aloud. 2) It could be a practice or facilitation effect. Asking people to produce two types of reports rather than one means they are asked to talk more; this may lead to an all round increase in talking.

The rate of verbal production is also relevant here. Group 2 respondents, who say slightly less, verbalise slightly slower than Group 3 respondents who say most and say it more quickly. This suggests that Group 3 respondents have slightly more fluency in reporting their thoughts.

Of the three think aloud groups then, the think aloud of Group 3 may by slightly more compete than the other groups. Added to this there is also the extra information often collected in the retrospections of this group, making the reports more complete.

The amount of information provided in the retrospective reports of Group 4 (retrospection only) compares favourably with the amount of think aloud produced by the other groups.

Thinking aloud generally increases the time taken to complete a task (Ericsson & Simon 1980,1984; Payne, Braunstein & Carroll 1978; Fidler 1983). This is because translating one's thoughts into words takes time. An increase in the time taken for completion does not, however, necessarily indicate a change in processing.

In this study Group 4 do not think aloud, but only retrospect. This is the closest condition to a 'response only' group in this experiment, and may provide a base rate against which to compare response time. Using this group as a base rate is to some extent problematic -- as noted above the instruction to retrospect may increase talking generally. Group 4 do produce

some verbalisation during response. This latter however, need not be seen as a problem. The 'natural' base rate should not be one of no think aloud but one which reflects what happens in a normal survey interview.

Whilst information on normal response behaviour in the survey interview is scarce, some evidence suggest that respondents do tend to verbalise more than simply a response. A study by Sykes and Morton-Williams (1987) uses interaction analysis, a method which as they say 'depend[s] on the outward manifestations of mental processes'. Behaviour from the respondent includes such things as questions, speculating on meaning, inappropriate responses, digressions, and qualifications of responses. Dijkstra, van der Veen, and van der Zouwen (1985) report similar respondent actions. These types of 'verbal reports' are similar to the verbalisations during response given by respondents in group 4. Thus, the base rate provided by group 4, while not perfect, is probably adequate.

Surprisingly, only for group 3 respondents is response time significantly slower than for group 4 respondents. This may simply indicate that group 4 does not present a good base rate for response time. If group 4 does provide an adequate base rate for response time, then this result may be echoing the suggestion above that group 3 respondents are reporting more fully than the other two think aloud groups, taking more time to translate thoughts into verbal code.

3.4.2 How valid are the protocols?

Problems with validity for think aloud protocols stem from the possibility of interference with the ongoing process. Problems with retrospective reporting derive from the respondent having an overview of the completed task and thus an opportunity to rethink it.

Byrne (1983) suggests that think aloud should be used only for those tasks on which respondents find it easy to think aloud. The reason being that, in these cases, think aloud is less likely to interfere with processing. From the results presented here - the fact that there is some verbalisation in group 4 and the fact that respondents tend to talk quickly - it seems that, while there may be individual variation, most respondents found thinking aloud both an easy and a natural task. Some respondents' post interview reports also support this.

The structure of the language can provide clues to the validity of the data, especialy for think aloud protocols. In the think aloud protocols produced here the structure of the language is consistent with what would be expected if respondents were reporting the contents of currently heeded thoughts (Ericsson and Simon 1984). Statements are generally in the present tense, or reports of information retrieved. There are few statements of what was "thought", and few descriptions of processes. Thus, think aloud protocols contain the information that was processed, not discriptions of how it was processed (Svenson 1989).

The language, while generally fluent, does not appear to be carefully planned, containing as it does a number of poor sentences from generally articulate speakers ie. the syntax also varies from normal speech (Bainbridge 1985). Respondents tend to begin verbalisation immediately and to talk relatively quickly. Thus, at this level, there is no evidence indicating that think aloud reports in all groups are **not** reports of currently heeded information.

Thus far, the results regarding the structure of the language, the time taken to verbalise, and the ease of verbalisation, all support the proposition that think aloud does not alter processing.

Retrospective protocols can be regarded as a less obtrusive measure than think aloud. In this case one might expect a comparison of the content of the retrospective protocols of group 4 with the think aloud groups' protocols to reveal any inadequacies with think aloud. If the protocols are similar then it can be argued that think aloud has not interfered, if different then it has. However, as Svenson (1979) points out retrospective reports require special care in interpretation because they provide the respondent both with sufficient time and with a completed process which may allow them to apply their own theory of the process. Thus, retrospective reports may be more open to the kind of criticisms of verbal reports voiced by Nisbett and Wilson (1977).

In fact the content of Group 4 respondents' retrospection is generally more similar to think aloud protocols than to the retrospective protocols of group 3. However, there are some notable worries about the retrospection of group 4.

Firstly the differences between the retrospective protocols of group 3 and group 4 seem mainly to be due to a difference in reporting style. A large amount of the differences between groups 3 and 4 in the content of their retretrospections derives from the use of process descriptions by group 3, but not by group 4. Also group 3 contains little direct repetition of what they said in the think aloud protocols, rarely is the actual response given whereas in group 4 this is more common.

The retrospection of group 3 respondents should not be interpreted without taking into account the context that they have already thought aloud. They have already said what they are thinking, in many cases, and to report again so quickly in the retrospection may be a violation of a conversational 'given new' contract (Grice 1975). In essence, respondents may think 'well she obviously doesn't want to hear what I've just said, because I've just said it, so now what else was I thinking about', or 'what was my thinking like'. This seems to be the case for one respondent who says for her retrospection 'And that was it, I didn't think of anything else, except what I said'.

Perhaps respondents think that they can't just repeat exactly what was said, so, they sum it up, elaborate on instances mentioned, or mention other thoughts or describe 'how' they thought. Group 4 respondents, not having the think aloud, do not have to do this and so report more statements.

Thus it may be that many of the differences between the retrospective reports is not due to think aloud changing processing, but rather that think aloud changes what is reported in the retrospective reports, or how it is reported. The face-to-face interview setting may enhance this effect.

Process descriptions can be problematic because it is difficult to maintain that these are reports of thoughts rather than theories about thinking. Though these kinds of statements occur in both groups, they are more common in group 3. However, when they corroborate statements given during think aloud, they become less problematic.

One factor which may influence the reporting of processes is the prompt used for retrospection -- "Now, please tell me all that you can remember about your thinking". This prompt might be interpreted as asking the respondent about the way they were thinking. A prompt such as "tell me what you were thinking about", may focus more on the content of thought. Having thought aloud this interpretation may provide an easy answer to the problem of redundancy for group 3 respondents. Thus, the prompt might have encouraged respondents to give their ideas about their thought processes.

In the content of the protocols there are some differences which cannot be accounted for by the differences in the style of the reporting by groups 3 and 4, and are thus more worrying, as is the one difference in amount of retrospection on question 4, where group 4 say substantially more than other groups. These differences suggest more a problem with retrospective protocols than a problem with think aloud.

Some of the problems in retrospective reports are evident in the following statements, both taken from group 4 respondents. Are these reports of what was thought while responding, or are they reports of ' the kind of things I must have thought about' ?

"I can remember just picturing certain things. I mean those sort of kind of images which come to mind. Um I mean images say of the catalogue " or
"so normally I go to the other ones, to the microfiche or the catalogue, no I've never been to the catalogue only the microfiche when I got problems. For example Um when there are some things you have to find at the I mean the main desk".

The difference in amount of retrospection for question 4 (WHAT USE) is difficult to explain. Theoretically, one could argue that this type of short, recall question should be more likely than other questions to produce automatic, or at least very quick processing, and, thus, one would expect very little think aloud and similarly little retrospection. This is what one finds in the think aloud and so this result would argue

for faults in group 4 retrospective reports rather than problems with the other groups.

Burton and Blair's (1991) view was that the short recall task they studied may be disrupted by a think aloud procedure making retrospective protocols superior. In contrast to this view the present study suggests that think-aloud may be more accurate than retrospective protocols. Perhaps in the retrospective reports respondents feel more pressure to report something, even if there was very little actually remembered thinking.

The other evidence (from the structure of the language in think aloud protocols, the rate of verbalisation, the fact that think aloud seems to be relatively easy for subjects) argues, that while there may be some minor problems, the think aloud procedure is valid for gathering information in this way. It seems, as others have found (eg Fidler 1983), that retrospective reports may occasionally be more problematic than think aloud reports, and that retrospective reports may be more reliable when given after think aloud than when not, because of a stronger memory trace, and through the correspondence between the two reports.

3.5 CONCLUSIONS

The results suggest that think aloud is likely to be a valid method for use with survey questionnaires. Of the instructions used here, instruction C (think aloud plus

retrospection), seems to be the best because it seems to encourage respondents to speak more. Problems were noted with retrospective reports. However, it should be noted that these can be very useful when used in conjunction with think aloud, because they allow further information on very brief processing.

The interpretation of the results suggests that some modifications be made to the instructions. Firstly, the prompt given for retrospection should be changed in order to discourage explanations of processing. Secondly, more time should be taken to ensure that respondents understand what they are meant to do. This could be in the nature of a discussion of the instructions with respondents as suggested by Svenson (1989). This might also help to deal with some of the variation between respondents.

One aspect that has not thus far been explicitly addressed is the amount and type of information that is **not** reported in think aloud. Here, the problem is that while what is said is obvious, there is often little evidence of what is not said. It is very likely that respondents are not reporting all their thoughts, partly because there are too many rapid thoughts to report. There is some direct evidence of this in the protocols, however, there are probably many more cases where no report is given because of automatic processing which goes unnoticed by the respondents.

For example, many processes involved in comprehending the question may be very quick and may begin before the question has been fully asked. Robertson, Black and Lehnert (1985) for example suggest that the first word of a question such as 'how' or 'when' already leads subject to access expectations of types of answers that will be possible. Also, if as Robertson et al suggest, question understanding is an integrated rather than a modular process then one might expect that, as well as some understanding processes, sources of knowledge have been accessed and pieces of information have been retrieved before the interviewer has finished asking the question. It is unlikely that these types of processes will be reported in think aloud protocols, and, indeed, there is little evidence for them from this study.

There are, though some very explicit attempts to comprehend the question. Some of these attempts do not come immediately after the question has been asked but at a later stage, suggesting, perhaps, there is an attempt to re-understand the question (Galambos and Black, 1985). These more thoughtful attempts to understand the question seem reportable. However, many of the quick, implicit attempts at understanding go unreported and perhaps do not leave their trace in further statements from which they could be inferred.

While think aloud protocols are useful for getting at the conscious attempts to answer questions, there are likely to be a number of processes that go on below the level of reportability (perhaps what Byrne 1983 refers to as the

'comment level'), leaving no evidence of their occurrence in the protocols. Thus, questions which require more conscious effort on the part of respondents seem likely to be most amenable for study with this method. Short recall questions, which require only a yes or no response, seem least amenable to this procedure. Other recall questions will vary depending on the amount of information or judgement needed by the respondent to answer.

This does not invalidate the use of verbal protocols, but suggests that, like most other kinds of data, they have their limitations and must be supplemented by other techniques aimed at addressing the types of information which think aloud does not provide. Given the relative recency of theorising on the processes underlying questionnaire response, think aloud has much to contribute since our knowledge, even of the more conscious processes of question answering, is limited.

CHAPTER 4. EXPERIMENT 2 - PART I.

COGNITIVE PROCESSING FOR BEHAVIORAL QUESTIONS

SUMMARY

Having determined in the previous study a suitable method for collecting verbal reports of cognitive processes in the survey, the present study then uses this method to gather data about these processes. Respondents verbalised their thoughts while responding to one of two versions of a questionnaire. The questionnaire contained both behavioural and attitudinal questions. In this chapter the results of the behavioral questions will be presented; the following chapter will deal with the attitudinal questions.

4.1 INTRODUCTION.

It is assumed that a variety of inferential strategies are used to answer behavioural questions. Yet, details of the use of various strategies have not been extensively investigated. Some studies have provided evidence of the use of inference strategies with behavioural questions (Blair & Burton, 1987; Burton & Blair, 1991); these have been discussed in chapter 1. The present study aims to extend this work by examining the strategies used when answering a number of behavioural questions, using verbal report techniques. Thus a general

aim of this study is to examine the types of processes used by respondents to answer behavioural questions.

As well as being useful for explorations of cognitive processes, verbal reports can also be used for hypothesis testing (Ericsson & Simon, 1984). The present study uses them for both purposes. Two particular hypotheses regarding the way in which response scales can provide information to respondents are investigated. Specifically these are the 'comparison shift' hypothesis and the 'meaning shift' hypothesis. Both of these hypotheses are discussed in chapter 1 (pages 40 - 42) but I will give a brief recap of them here.

The comparison shift hypotheses states that response scales at a previous question can influence responses to later questions by providing information to respondents about how they compare with others (Schwarz, Hippler, Deutsch & Strack, 1985). The scale is seen as providing a distribution in which people can place themselves. This type of informational influence may be most likely with questions concerning unambiguous behaviours. The meaning shift hypothesis states that scales provide information to respondents about the meaning of a question, and is likely to operate when the behaviour in question is ambiguous (Schwarz, Strack, Muller, & Chassein, 1988). Priming the area has been shown accentuate the meaning shift (O'Muircheartaigh, Gaskell, & Wright, 1992).

These hypotheses have previously been investigated through split ballot techniques (O'Muircheartaigh et al, 1992; Schwarz

156

et al 1985). The aim of the present study is to look for evidence of the hypothesised underlying processes in verbal protocols. Thus, for the comparison shift hypothesis the protocols should provide evidence of people using the scale to locate themselves in comparison to others. To test this hypothesis an unambiguous behaviour, concerning the amount of coffee drunk is used. For the meaning shift hypothesis the protocols should provide evidence of people using the scale to interpret the meaning of the question. To test this a question about annoyance with television advertising is used, an ambiguous behaviour.

Thus the protocols should provide evidence of the cognitive processes used by respondents to answer behavioural questions. If the comparison shift and meaning shift hypotheses are correct, these processes might be expected to include references to the scale either as a comparison point, in the former case, or as a means of interpreting the behaviour in the latter case.

4.2 METHOD

Questions were varied between experimental conditions. Each group answered one of four 11-item questionnaires. Respondents were instructed to think-aloud while responding and to recall their thinking after each question. Questionnaires were administered in a face-to-face format. Three different interviewers were involved. Respondents then completed a short computer task which will be described in the

next chapter. The first four questions were behavioral questions, the next three questions measured relationship and life satisfaction, whilst the final four questions were attitude questions (the questionnaires are given in appendix 5). In this chapter the four behavioral questions will be discussed.

With one exception these questions are drawn from a survey study by O'Muircheartaigh, Gaskell, and Wright (1992; Gaskell, Wright, & O'Muircheartaigh, 1992), allowing for the laboratory testing of field generated hypotheses, and comparisons between think aloud and known survey findings. The exception is the priming questions used (TVHOURS and CLASSHOURS) to test the meaning shift hypothesis. The original study showed that priming increased the meaning shift effect and was therefore an important factor in looking at the meaning shift hypothesis. However, the original study had a number of priming questions which would have been too large for the present think aloud study, and so only one general question was used.

4.2.1 The respondents

Forty-one men and women were recruited from coffee bars and classes around the LSE. Most were students. Their ages ranged from about 18 to 50, with most in their twenties.

4.2.2 Procedure

In the same way as for the previous experiment, respondents consented to be recorded and then read the instructions on

think aloud and immediate retrospection. These instructions were the same as condition C in experiment 1 (see appendix 1). A minor change was made in that the prompt to retrospect was changed to 'Now tell me all you can remember thinking about'.

Once respondents had finished reading the instructions, they were asked to tell the experimenter briefly, in their own words, what they thought the instructions meant. Any misunderstandings were then clarified by the interviewer. Again, two practice questions were then administered the first being the same as in experiment one. The second, was a more attitudinal type question which was thought to more clearly represent the balance of the questionnaire to follow.

During the practice session the interviewer attempted to give feedback on whether the think aloud and retrospection was being done properly. Thus, there were more attempts than in experiment one to clarify with the subject how to think aloud and retrospect. The questionnaire was then administered. Finally, after completing the questionnaire respondents were asked to complete a short task on the computer which was to compare items for similarity / dissimilarity. Respondents were paid for taking part.

4.3 RESULTS

Each question, or set of questions, can be seen as a separate experiment. Questions which address different hypotheses will

be grouped together, giving an introduction to the questions and the hypothesis, the results and a discussion of how the results address the particular hypothesis.

4.3.1 The comparison shift hypothesis.

Two questions are involved in looking at the comparison shift hypotheses.

The questions

The first question (AMOUNT COFFEE) asked respondents to report the amount of coffee they usually drink in a day. This represents a mundane, unambiguous behaviour, and asks for usual behaviour. There were two versions of this question. One group received a High Frequency response scale and the other received a low frequency response scale (see figure 4.1 below). A response card with precoded categories was used for response. It was expected that there would be little differences in either responses or strategies to this question as a function of the scale; because the behaviour is unambiguous the respondents should not look to the response scales to provide an interpretation of the behaviour.

Figure 4.1
COFFEE QUESTION AND RESPONSE SCALES

Q. How many cups or mugs of coffee or tea do you usually drink in a day?	
Response Scales	
High Frequency	Low Frequency
16 or more	4 or more
10-15	3
4-9	2
2-3	1
1	less than 1
never	never

The second question (COMPARE) asked people to estimate how much coffee they drink compared to others. For this question, we might expect to find differences between those who have been presented with a high frequency scale at question 1 and those presented with a low frequency scale. In particular, those given a low frequency scale should see themselves as drinking more coffee than those presented with a high frequency scale. That is here we would expect a 'comparison shift'. This question was the same for all respondents, and is given in figure 4.2 below.

Figure 4.2
COMPARE QUESTION

Q. Taking your answer from this scale, compared to other people, how often would you say that you have coffee or tea to drink?
Much more than most
A bit more than most
About average
A bit less than most
Much less than most

In terms of the protocols the comparison shift hypothesis suggests that people will refer to the scale at the previous question, AMOUNT COFFEE, to provide an estimate of what an 'average' coffee drinker is and how they fit into that distribution.

Results

Question 1 - AMOUNT COFFEE

Responses

The distribution of responses between groups is given below (Table 4.1). Categories have been combined to make comparisons possible.

TABLE 4.1
Responses to COFFEE

	High Freq	Low Freq	TOTAL
4 or more	17	12	29
2-3 cups	3	4	7
1 or less	1	4	5
TOTAL	21	20	41

There are only minor differences in response across the groups. This is expected. As the question is unambiguous as to the type of behaviour required, differences in the response scale should not influence choice. In a sample survey using the same question Gaskell, Wright & O'Muircheartaigh (1992) found similar results.

Protocols.

In terms of describing the processes used to answer this question, the main strategies used for response were identified. For question 1 two main strategies can be discerned. The first is to log through the day and count the number of times coffee or tea was drunk (TimesDay). The second is simply to make a direct estimate, to grab a number. This can either be a response from the scale or not. Table 4.2 summarises the main response strategies breaking them down across groups.

TABLE 4.2
Response Strategy for COFFEE by group.

	Hi Freq	Lo Freq	TOTAL
Timesday	12	8	20
Direct Estimate	5	4	9
Response	4	4	8
Other	0	4	4
TOTAL	21	20	41

Respondents who use a TimesDay strategy vary in how completely they use this strategy and quite how they link it in with a day. Half of all respondents use some form of this strategy. There is a small difference in the amount this strategy is used across conditions. More respondents use it in the high frequency than in the low frequency group. This appears to be partly due to the fact that for some respondents in the low frequency group the scale 'provided' the answer for them; that is, since 4 was the highest category, they did not need to count the number.

An example of a strong use of this strategy is given by respondent 56: "I usually have about - one at coffee, one at lunch, one in the afternoon, a cup in the evening....".

The day tends to be divided into segments of morning, afternoon and evening; or breakfast, lunch and dinner; or simply 'times I stop for a break'. The number of cups in each segment is counted and added, though this is not usually explicitly stated in the protocol.

This strategy is a decomposition strategy, perhaps reflecting a general day schema, looking for 'natural' breaks in the day when coffee or tea is likely to be drunk. There is very little recall of specific events, rather there is recall of generalities.

17 respondents make a direct estimate, of these 9 give categories from the response scale while the other 8 give a response. It is, of course, possible that the people adopting this 'strategy' are simply not reporting all their thinking. However, many of these people do say quite a bit, none of which reveals counting through the day.

The above represent general strategies, but a number of other processes are also evident in the protocols. Some are peripheral, whilst others seem to contribute to the general strategy.

Location indicators play a part in both strategies (kitchens, cafés), whilst time indicators are restricted to the Timesday strategy. For a number of respondents imagery plays a role, with images of cups and kettles being common.

Very few respondents mention difficulties in recall or estimation, and there is very little change of strategy. Only one respondent reports thinking about what drinks to include. Thus, for the majority of respondents, it seems the question is taken to be relatively straightforward and unambiguous.

QUESTION 2 - COMPARE

Responses

The breakdown of responses to question 2 across groups receiving different frequency scales at question 1, COFFEE, is given below.

TABLE 4.3
Responses to COMPARE

	High Freq at Q1	Low Freq at Q1	Total
much more	2	1	3
a bit more	6	6	12
about average	8	4	12
a bit less	5	5	10
much less		3	3
dk		1	1
TOTAL	21	20	41

There are only slight differences in the distribution of responses with this data. Using the same questions in a large

scale survey, Gaskell, Wright & O'Muircheartaigh (1992), found that those given the low frequency scale were more likely to say they drank more than most than those given the high frequency scale. The data in table 4.3 above do not conform to this pattern, however, because of the small sample size, and the resultant small numbers giving extreme responses, this response data is less reliable than the larger sample used by Gaskell et al.

Protocols.

Firstly the protocols were examined for reference to the previous scales and for estimates of what average is. No references to the previous scale were found and very few people indicated what they thought average coffee consumption was. General strategies were then compared.

The strategy used by most respondents is to compare themselves to others who they know (KnownOther). For example one respondent reports " I compared myself to my family, who, that's who I usually sit and have my drinks with", or ".. I compared myself to her because she's the only person I know who drinks a lot of coffee".

There are a number of other, less commonly used, strategies. The second most common strategy is thinking about what most people, or what an average person, would drink. Usually there is no clear indication of what is 'average', but rather a more vague sense of what most people do. Those who base their comparisons against 'most people' do not go into great detail

working out what 'most people' or an average person would drink. For example "It just occurred to me what people , what an average person would drink. And, who didn't like coffee and who likes lots. And, it sort of equals it out in the end." Seven people use this strategy. Another 2 are unsure of their response because they don't know, or are unable to generate, an estimate of what most people drink, though they attempt to do so.

Other strategies include five respondents who say that other people tell them that they, the respondent, drink a lot or a little; two respondents who 'Just Know' how they compare to others; four respondents who give only a response; and three who use other strategies.

TABLE 4.4
Response strategies for COMPARE

	High Frequency	Low Frequency	Total
KnownOther	12	6	18
Most - Average	4	5	9
PeopleSay	3	2	5
Justknow	0	2	2
Response	0	4	4
Other	2	1	3
Total	21	20	41

Thus, 27 respondents attempt in some way to compare themselves to others, or to think about how much others drink, while 14

do not. However, most people are not comparing themselves to an average, but simply to people they know.

These strategies vary across conditions. Those who receive the high frequency scale with question 1 tend to compare themselves to people they know more than those who receive a low frequency scale who tend, on the other hand, to respond somewhat more superficially.

Discussion of comparison shift

Although there is no direct evidence to support a comparison shift (that people see the scale as representing a distribution of the behaviour) there are differences in the strategies used between the groups. This suggests that the scales affect response, but perhaps do so in a more subtle way.

Quite why one scale, the high frequency scale, should lead to more comparisons with others is uncertain. One possibility is that the high frequency scale alerts respondents to the range of behaviour, thereby making some kind of comparison more salient. Perhaps a more precise scale deserves a more precise answer. It is notable that the high and low frequency scales used here, although having the same number of categories, have very different ranges within the scale, taking the top mentioned number as an end point the ranges are 16 and 4 points. Thus, it may be simply the breadth of the range, rather than its particular end points, which is influential. It would be useful to compare scales with the same breadth but

different end points. Given that a large range is made salient, the comparative aspect of the question may become more salient. The tendency for some people to compare themselves with 'coffee drinkers' may account for why more people see themselves as drinking less coffee at the high frequency scale according to the comparison shift hypothesis.

Another possibility comes from work done by Martin & Harlow (1992). Looking at the effects of filter questions, they suggest that answering filter questions, questions such as who one's political representative is, can lead to affirmation or disconfirmation of self-concepts. The immediate effect of this on thinking, they suggest, is that those who are successful in answering such questions are likely to bask in success, thinking of the issues related to their successful confirmation. Those who are unsuccessful, receiving a blow to their self-concept, are likely to attempt to think about something else, to distract themselves from their failure. And indeed, they provide evidence to support this hypothesis.

Perhaps this disconfirmation of the self extends to other types of manipulations which threaten the self. If when placing themselves on the scale people do treat it as representing the distribution of coffee drinkers, then those answering on the low frequency scale would appear to drink a lot of coffee. This may arouse anxiety; there is evidence in the protocols that people are worried about drinking too much coffee. In this case people may not want to think about the subject which arouses anxiety and so give more superficial

responses, not deliberately comparing themselves to others.
The response may then be drawn, subconsciously, from the
implicit judgement made at the previous question.

4.3.2 The meaning shift hypothesis

In this study the hypothesis was assessed by the use of an
ambiguous behaviour -- frequency of annoyance with television
adverts. The question received a 2 by 2 treatment. It was
either primed or not primed. Thus half the respondents first
received TVHOURS asking for the amount of TV watched last
week, which serves to prime television viewing. The other
half received the question CLASSHOURS, asking for the amount
of time spent in lectures and classes last week, which acts as
a no-prime filler question. Both questions have open
responses. These questions are present in figure 4.3 below.

A question asking for the frequency of annoyance with
television adverts (ANNOY) then follows. This represents a
vague (and subjective) behavioral question. In each group
(prime vs no-prime) half the respondents receive a high
frequency scale and half receive a low frequency scale. It is
expected that this question would produce a 'meaning shift',
that is, respondents will use the scale to interpret the
meaning of annoyance with an advert. The questions and
response scales are given below.

FIGURE 4.3
Questions and Response Scales for testing 'Meaning
Shift' hypothesis.

PRIME	NO PRIME
TVHOURS Q3. Last week how many hours did you spend watching television?	CLASSHOURS Q3. Last week how many hours did you spend in lectures and classes?
Q4. How often do you feel annoyed by an advert or commercial on television? (ANNOY)	
High Frequency Everyday Most days Once a week Once a month Less often Never	**Low Frequency** Once a month or more Once every few months Once every six months Once a year Less often Never

QUESTION 3A - TVHOURS

Response

The mean number of hours television watched is 5.6 with a
standard deviation of 5.9.

Protocols

For this question, it is difficult to encapsulate the types of
strategies used. Partly, this is because many subjects use a
variety of strategies to respond. A table of strategies is
presented below.

171

TABLE 4.5
Strategies for TVHOURS

GENERAL STRATEGY	FREQUENCY
General activity + other methods	8
Attempts to recall/thinks of what watched	4
Concentrates on normal viewing	2
Direct estimate	3
General estimates	3
Other	1
TOTAL	21

A number of respondents begin by trying to recall their general activities last week; some then explicitly infer the general amount of TV viewed on this basis. For example, one respondent says " I think I spent a lot of time writing essays and preparing seminars so I actually spent very little time watching TV, maybe one hour a night, that would cover news programmes mainly." They may then go on to recall a specific, vivid programme watched, general programmes watched, or try, but fail, to recall what they have watched. In the latter case, those respondents may then go on to relate last week to their 'normal' television viewing.

There are also a number of minor strategies. Firstly, 3 people who rarely watch or don't own a TV give a direct estimate ('none'). Three attempt to recall what they have watched but fail and resort to a number of estimation strategies, including guessing. One, knowing he can't recall what he saw, simply guesses. Two think of regularly watched programmes and calculate from this -- one taking into account other vivid programmes, the other not doing so. Three use

general estimates of amount for different time periods. One uses an idiosyncratic strategy.

As can be seen, not only are there a number of general (overarching) strategies, but, in most cases, these are not pure strategies. Respondents are often using a number of different recall and estimation strategies to produce an answer, usually at a fairly superficial level, often little more than a guess.

Many people have difficulty in thinking how to answer the question, as evidenced by changes in strategy in the course of responding; when one avenue fails, they try something else. Many are aware of the difficulties and a number of respondents state that they are uncertain about their answers or their memories or that they find it difficult to recall their behaviour last week.

QUESTION 3b - CLASSHOURS

Response

The mean number of hours of lectures attended was 8.1 with a standard deviation of 5.3.

Protocols

For Classhours, (n=20), most respondents use some kind of rule based estimation to calculate an answer. Either starting from their normal hours and adjusting, or multiplying the number of courses by the number of hours or days. Only 2 respondent use (or try to use) episode enumeration.

TABLE 4.6
Strategies for CLASSHOURS

STRATEGY	FREQUENCY
Number of Classes X Hours	7
Total hours + adjustment	5
Episode Enumeration	2
Direct Estimate	4
Other	2
TOTAL	20

It is notable that calculating is virtually the only thinking going on. Very few people say, and thus think, of anything but the rule and the calculation. Very few recall either specific incidents last week or recall in general any imagery or otherwise surrounding their lectures and classes.

Of those who start from normal hours and then subtract some amount from that, most do not recall specifically missing classes or lectures but subtract because 'I usually don't attend them all'. Also there are few shifts in strategy. Most people stick to the strategy they first come upon. In this way there is less variation in the thinking behind answering this question than that involved in other questions. There is less reliance on individual episodic memories and more on rules. This is just what one would expect from this kind of question, that is, one which asks about a regularly occurring activity.

QUESTION 4 - ANNOYED WITH TELEVISION.

Responses

The responses to the questions are given below. Categories were collapsed for comparison across conditions.

Table 4.7
Responses to ANNOYED by group

	Prime High	Prime Low	No-Prime High	No-Prime Low
Once month +	6	8	5	7
Less Often	4	3	2	2
Never			3	1

There are only minor differences between the groups. Using this same question in a large scale survey O'Muircheartaigh et al (1992) found a small, but significant increase in the number of people reporting annoyance once a month or more for the high frequency scale.

Protocols.

First the protocols were examined for references to the scale, however, none were found. Questions were coded to determine the main strategies used by respondents. Table 4.8 below gives a breakdown of these strategies across the four groups.

Table 4.8
Strategies for ANNOYED by group

	Prime High	Prime Low	No Prime High	No Prime Low	Total
Exemplar (Ad or Annoyance)	1	7	1	2	11
Weak Exemplar	3	3	1	2	9
Inferential	5	0	1	1	7
Mix-Exemplar + Inference	0	0	2	1	3
Superficial	1	1	5	4	11
Total	10	11	10	10	41

All these strategies represent the respondents main strategy. The mixed strategy is one where two strategies appear to be used equally.

The 'Exemplar' strategy is where the respondent reports either thinking of a specific annoying advert or type of annoying advert or where they try to determine what is an annoying experience. For example, "I'm thinking about all those soap powder ads", or, another respondent described a greeting card that annoyed her and explained: "I just thought of something that really annoyed me, that made me angry", and then related this experience to adverts. Those coded as weak exemplars are respondents who simply stated they thought of 'various adverts' or where they tried but failed to think of an annoying advert.

The inferential strategy is one where the respondent bases their estimate of annoyance on the amount of TV they watch; for example, "I don't watch tele everyday and see adverts so, probably less than once a month". The mixed strategy is a combination of this and the above exemplar strategy.

Strategies were coded as superficial if they consisted either of a simple response, or if the respondent simply reported their general attitude, or some other general statement (eg 'I'm a fairly tolerant person').

The strategies are differently distributed across the groups. Looking first at the differences between primed and unprimed groups, the major difference seems to be that while in both primed groups there appears to be a dominant strategy (one used by at least half the group), in the no prime groups there is no dominant strategy. In general, the respondents in the no prime groups reply more superficially.

In the primed groups the strategies respondents use differ across response scales. More respondents in the primed high frequency group base their responses on the amount of television watched than do respondents in the primed low frequency group. This latter group think of an exemplar (or try to think of an exemplar). Generally, there is more thinking in these protocols than for the primed high frequency group. All the respondents who use an inferential strategy in the primed high frequency group use only this strategy; they mention nothing else.

It is also notable that the low frequency groups (both Prime and No Prime) are the only ones where some respondents explicitly attempt to define annoyed, or raise questions or doubts about what annoyed means. Seven of these respondents explicitly question, or try to define, annoyed or annoyance with adverts in some way.

Discussion - meaning shift

As with the comparison shift hypothesis there was no direct evidence for the meaning shift hypothesis. However, there was a difference in the strategies used between groups.

The results here are generally in accord with the meaning shift hypothesis, but qualify it somewhat. According to this hypothesis what might be expected in the protocols is reference to the scale during attempts to understand the question. In fact, there are no such explicit references to the scale. Instead, there are strategies which attempt to understand what annoyed is or which try to generate annoying instances. According to meaning shift, we might expect these strategies to be equally apparent in both groups, but, to involve different examples of annoyance between groups, unusual or more severe ones for low frequency respondents, trivial ones for high frequency respondents. But this is not what occurs.

It seems that what may be happening in this study is, firstly, that a prime focuses respondents onto a relevant strategy. The strategy used depends upon the response scale. Thus, for example, for a high frequency scale the amount of TV viewing may be

relevant (if you do not watch TV everyday you can't be annoyed at TV adverts everyday). For the low frequency scale the response categories do not suggest a need to look at the amount of TV watched as a priority (since most people will watch TV at least once a month). It could be that a low frequency scale prompts a strategy which concentrates on the other salient features of the question such as annoyance and adverts.

Secondly, there is a suggestion that the interpretation of 'annoyance' depends on the scale received. One of two things seem to be possible when the respondent hears the question: either they form an interpretation of annoyance or they register it as ambiguous. Upon receiving the high frequency scale their interpretation is either confirmed or their doubts cleared up. It may be that for the high frequency scale the interpretation of annoyance is made unconsciously, perhaps because that type of frequency scale fits with their everyday definitions of annoyance in the context of advertisements. Thus there is no explicit decision about what annoyance is and not much thought about what an annoying advert is. However, in the low frequency group, for some respondents, the scale may either clash with their idea of what annoyed means or fail to easily clarify what annoyed means, thereby creating uncertainty over the behaviour and a need to think about what annoyed might mean. These respondents attempt either to define annoyed in some way or to generate instances of annoying adverts.

In fact this interpretation fits quite will with the results found by Schwarz et al (1988) when they asked respondents how

often they were annoyed. As well as the two conditions where a frequency scale was presented (high or low) another condition had an open response format. After responding to the question respondents were asked to report a typical example of the annoying experiences they had had. These examples were then rated for severity and concreteness. Those given the high frequency scale and those given the open response format reported examples of a similar level of severity; those given the low frequency scale reported more severe examples of annoyance. This fits with the present suggestion that the high frequency scale fits more with the respondents definition of annoyance. Furthermore those given the low frequency scale reported more concrete examples than those given the other reporting formats. Schwarz et al suggest that this indicates that the more annoying experiences were represented in memory in more detail. The present results suggest this may be because in the case of the low frequency scale respondents had previously generated examples while responding, whereas those in the high frequency scale had not.

Together these results suggest that when scales do not fit with a more normal understanding of a term, respondents may engage in attempts to understand what the behaviour is. This may or may not push them in the direction of defining the behaviour as more severe (or less depending on which scale causes ambiguity) depending on how easy it is to generate examples of this type of extremity, which in turn probably depends on their knowledge of the behaviour. With the questions used by Gaskell et al (1992) the behaviours queried are less ambiguous than that used by

Schwarz et al; annoyance with television adverts produced the greatest response effect. It may be for these behaviours that, although respondents consider the meaning of the scale that does not fit, they are unable to generate more severe example.

4.4 DISCUSSION

Inference strategies and accuracy

In this section I want to discuss more generally how people have responded to the behavioural questions: what types of processes occur, how prevalent their occurrence is and whether they differ across question types.

The first thing to note is the diversity of processing both across questions and, for most, within questions. The questions used call for different types of behavioural estimates: estimates of number in a specific period (CLASSHOURS and TVHOURS), the rate of occurrence (ANNOY), usual behaviour (COFFEE), behaviour compared to others (COMPARE). In part this diversity may be due to the different types of estimates required.

Firstly there is extremely little use of episode enumeration strategies. This strategy might be most expected at CLASSHOURS and TVHOURS which both ask how much the behaviour occurred in a specific period. From the literature on survey methods and cognitive theory however, these questions might not be expected to produce this type of strategy. CLASSHOURS asked for an estimate of a regularly occurring activity and TVHOURS for an

estimate of a relatively mundane activity. For both of these activities particular episodes might be expected to be blurred, and respondents are more likely to report on the basis of generic memory or memory schemas. Smith, Jobe and Mingay (1991) report use of generic memory for dietary recall, also a frequent and perhaps repetitive event.

For TVHOURS a number of subjects actually approach the question in a way that suggests they want to use a type of episode enumeration, but find it too difficult. One respondent says "I'm trying to remember. I like to think that I only watch TV for specific programmes and therefore I should be able to reconstruct quite accurately if I could remember the programmes, but of course I can't" . Thus other inference strategies are used, most commonly inferring from their general activities in the period and from their normal TV habits, occasionally decomposition by type of programme or by time of week (weekday or weekend) is used, but only minimally.

For CLASSHOURS only two respondents approach the question by looking at last week. Most respondents base their estimate on their normal behaviour. Inference strategies include direct estimates, anchoring and adjustment and rate based calculations from normal behaviour.

Thus for a regularly occurring behaviour like CLASSHOURS we might expect respondents simply to go by their normal behaviour. These types of regularly occurring activities are often important in surveys; number of hours worked in a reference period is an

important question on labour force surveys for example. The worry with this type of data is that it will be biased towards normal behaviour, even many of the adjustments made are based on normal practices. Abelson, Loftus, and Greenwald (1992) shows how habit can bias reports of voting. Habitual voters overreport the behaviour whilst habitual non-voters underreport. The same kind of reporting may occur here. Those who usually attend classes may overreport last week's attendance, whilst those who usually don't attend everything may underreport. The protocols make clear why this might happen. Indeed, at least two respondents indicate after giving their retrospective reports that they have given an incorrect response because they recall an extra or a missed lecture.

Precise estimates of mundane activities seem most difficult for respondents. TVHOURS and to a lesser extent ANNOY produce the most expressions of difficulty from respondents. ANNOY seems to be difficult because people aren't quite sure what it is, whereas TVHOURS is simply a difficult memory problem. One has to question the value of obtaining precise numerical estimates from respondents on this type of behaviour, since the responses given are clearly not precise. A number of respondents admit they are simply guessing whilst others indicate they are not being precise. The danger in collecting this type of report is in treating the numbers as more precise than they actually are. Much more intensive questioning would be required to produce estimates which are more reliable.

It is also worth noting that very few people define the reference period they are using. They seem to make an assumption about what last week is. This may vary across respondents.

Decomposition type strategies are most common for COFFEE. A number of decomposition type strategies are used, sometimes in combination. The most common is a time of day based decomposition. Although, as for most questions, respondents are not too careful in their estimates, in some ways these appear to be reasonable estimates of usual behaviour. However, there are some elements which may lead to bias.

The type of decomposition used focuses on " the number of times I stopped for a drink". There is little decomposition based on types of drink, so it is possible that respondents may be including other types of drinks. One respondent doesn't seem to think it that important which drink is consumed: " I'm not sure whether it's coffee or tea but it's a hot drink". Also there is little questionning of what a 'usual' day is. Only one respondent differentiates between weekdays and weekends, and one between college and home days. Most take a college day to be a 'usual' day. This is in fact when the interview takes place, so this may represent a bias towards the type of day on which collection takes place, or it may simply be a reflection that this is a more typical day. Also, no one, for example, considers such things as seasonal variations in drinking habits, perhaps more cold drinks are drunk in the summer. This suggests that estimates will be biased towards the time of collection. In short whilst decomposition strategies are used, they perhaps only

partially decompose the problem, and do not consider other aspects.

The use of exemplars is common for both ANNOY and COMPARE. Exemplars may be seen as the use of an availability heuristic, one uses the most available pieces of information to form a judgement. Kahneman and Miller (1986) have extended the work on availability (Tversky and Kahneman 1973) with norm theory, which proposes a model of how exemplars are used in judgement. They propose that a norm is produced for a category by combining a set of retrieved representations, these representations are salient or accessible exemplars.

Smith and Zarate (1992) also propose a model of social judgement based on exemplars rather than algebraic or schematic models. Rather than comparing to a prototype people may compare to exemplars; the exemplar does not have to be a particularly good example of a category to be used for comparison.

The accessibility of exemplars is obvious in COMPARE. Usually people compare themselves to friends or family -- people they are regularly with. Even when comparing more generally, without specific people, respondents generally compare themselves with their group - students. Few people take into account other types of people. One person recognises this but says "I have to decide what the other people are for myself, and I think it's fair enough to to just think of say you know my friends that I live with as as typical human beings." Thus, thinking of people one

knows simplifies the response problem. But it also shows how people can think that they, or their group, are the norm.

The problems associated with availability are apparent here. The people one knows may not necessarily form a good basis for comparison. Estimates from different subgroups may produce different estimates if these groups differ in the amount they engage in the behaviour. A heavy coffee drinker, surrounded by other heavy coffee drinkers may report they drink an average amount of coffee, but so too would a light coffee drinker surrounded by other light coffee drinkers. Wright, Gaskell and O'Muircheartaigh (1994) showed how both the individuals own behaviour, and the social group to which they belonged effected estimates of how much television a typical person watched.

There are differences in the number of exemplars used. Sometimes just one exemplar is used. One person compares herself to only one friend, or only one particular annoying advert is recalled. Others recall more. One respondent mentions two friends, one who drinks a lot and one who doesn't drink any, and puts herself in the middle. This example shows how the person used two extreme examples to arrive at a norm for coffee drinking. This recalling of both extremes is unusual however. Mostly people generate examples of people who engage in the activity or of *annoying* adverts. Those who like adverts don't generate examples of good adverts. The focus is very much on instances of the behaviour rather than contrary examples.

This tendency to generate confirmatory examples can also lead to biases. Concentrating on annoying ads, without considering all the ads that aren't annoying, may lead one to consider that the incidence of annoying ads is greater than it actually is.

Thinking of instances, rather than of generalities, can be found in other statements. People remember particular conversations or particular articles they have read which relate to the question.

Some people, especially for ANNOY, base their response not on the behaviour but on their general attitudes. At ANNOY it is their attitude towards advertising which may be the basis of the response. In some cases the attitude can completely dominate the response to the extent that one's estimate of the amount of annoyance bears no relation to the amount of television watched and thus to actual behaviour. For example, one respondent says she is annoyed by an advert everyday, previously she had reported not watching television. Another respondent says "I immediately thought my big thing is about how subversive ads are, and I was thinking gosh it's subversive. And if she gives me anything that says every 5 minutes then that'll be the one I'll choose. I liked the question, I thought that was a good question cos it's one I can respond strongly to." In these cases it seems the strict accuracy of the response to the question as posed is not what is centrally important to the respondents. What is important is getting their strongly held views across. This provides a nice example of the way in which the survey functions

as a communicative event, and illustrates how respondents interactive goals can influence response accuracy.

Sometimes the response is based on ideas about self, or theories of self. Here the respondent thinks about what they are like, and it seems implicitly infers from this how often they would engage in the behaviour. This is an estimation based not on retrieved events but on what a person like me would do.

Lay theories of memory also play a minor role. For ANNOY when people fail to generate an exemplar they may rely on implicit ideas about memory to judge what this failure means. Either deciding that because they can't recall, they can't be annoyed very often, or just because they can't recall doesn't mean they aren't annoyed.

In the protocols then we can observe a number of strategies which have been suggested to occur based on the application of cognitive theory to survey methods. A number of types of inference strategies occur based on comparisons with normal behaviour, anchoring and adjustment, exemplar based strategies and to a lesser extent theories of the self and memory. Some of the ways in which these processes may lead to errors was also discussed.

These processes reported above are, of course, all conscious. It is almost certainly the case that other processes, perhaps more of the same, perhaps different processes, have occurred which have not been reported. Some of these seem to be implied

by the protocols, but others probably are not. Four general types of processes are assumed to occur in response to survey questions (Tourangeau, 1984). These are comprehension, retrieval, judgement and response. Most of the types reported above would probably be seen as either retrieval or judgement. Comprehension and response processes are less represented. I want to look briefly at these processes and some other minor processes not represented in this four stage scheme.

Comprehension, Response and Other minor processes

There are few explicit statements about comprehension. Attempts to define what annoyed means generates the most comprehension problems. This is a vague term and thus is what one might expect. However, rather than simply occurring at the beginning of a response, these tend to occur later in the protocols, or throughout them. Some respondents focus almost exclusively on trying to understand what an annoying behaviour in relation to an advert would be.

Over all questions only a couple of respondents report any kind of immediate understanding of a question. It seems that, for the most part, comprehension occurs as a relatively automatic process, leaving little trace in memory. Comprehension processes may only become conscious when there are problems of comprehension.

As well as these kinds of definitional comprehension - understanding what the question means - there is also evidence of a more procedural kind of comprehension. This is more

understanding the question, and how to respond in a more general way, including requests for repeats of the question or specific items within the question as well as what they are meant to do with the response card provided. These types of statements occur for all questions but somewhat more for ANNOY.

Responding, as with comprehension, seems to proceed largely automatically. Mapping the response onto the response scale, produces little information from respondents. A few respondents indicate that they are looking for their formulated response on the scale. There are some indications of satisficing principles at work in response to both amount coffee and annoy. A few respondents pick a category because they 'like it' and a few also exclude categories. Some respondents at COFFEE indicate that the scale has provided them with the response, that it curtailed their need to think.

There is also evidence of response anxieties from a few respondents at amount coffee and at TV hours. However, this does not appear to be necessarily associated with a response stage, but rather seems to be a more general emotional reaction to the question. For one respondent their anxiety about drinking too much coffee dominates the response, and perhaps interferes with response.

Other minor processes also occur. There is some evidence of respondents setting up retrieval cues. These tend to be very general rather than specific. Some respondents indicate their level of certainty of the estimates provided, and estimate how difficult the question is. Largely, this is when they are

uncertain or when the think the question is difficult, although some indicate when they are certain of their response. There are very few judgements about the accuracy of a retrieved piece of information, and very few decisions to discard retrieved information.

Strategy decisions are also sometimes stated in protocols, a kind of meta-stage decision on how to go about answering the question, or even whether to answer it. Only a few respondents actually mention these strategy decisions, and they tend to come early on in the protocol, although there are some points where explicit strategy changes are made later in the protocols. Another example is the two respondents who decide whether or not to respond to the question. Although few respondents explicitly mention these procedures, it is possible that they are more common but more automatic.

4.5 CONCLUSIONS

Firstly, the protocols reveal that, in general, respondents tend to approach the questions in a more or less conversational way. Most do not really try to be exactly accurate. They guess, give rough estimates and rarely ever check recalled information for accuracy. They rely on a number of inference strategies which may lead to a number of types of biases in reports.

Secondly, in regards to the question of accuracy of verbal reports, the implications of these results are positive. The protocols differed across conditions as expected. Thus, as

Burton and Blair (1991) have found, experimental manipulations can lead to differences in verbal reports. By using verbal reports in an experimental setting it is thus possible to further look at their accuracy. Also, for some questions the types of strategies which might be expected to occur were found in the protocols.

In regard to the completeness of verbal reports, it is almost certainly the case that verbal reports do not capture all the processes which are occurring, and, especially for the relatively quick comprehension and response processes, other methods will be needed to better understand these processes. However, the information that is provided is relevant to the investigation of cognitive processing in the survey. Not only can it provide information on particular hypotheses, but it has been useful in providing information on a number of other types of processes.

Finally, the protocols have provided further information about the specific hypotheses tested. It should also be noted that the lack of evidence for some conscious processing is also interesting here. Where Schwarz and colleagues had predicted a 'meaning shift' and a 'comparison shift', although processing is altered, there is no reference to the response scales in the way which Schwarz and colleagures hypothesised. This suggests that while these factors are influential, they either do not influence thinking in quite the way hypothesised, or they do not do so at a conscious level. That is, even with a sophisticated sample, respondents do not consciously consider the response distribution as a population distribution, nor do they make conscious

decisions about the meaning which a scale provides. The way
these hypotheses have been stated in the past, one might have
expected respondents to do so. Instead, at least at a conscious
level, different scales seem to influence the strategies used for
response.

CHAPTER 5 - EXPERIMENT 2 --PART 2

VERBAL REPORTS AND ATTITUDE QUESTIONS.

SUMMARY

This chapter reports the results of the attitude questions asked as part of the experiment described in the previous chapter. A number of attitude issues were addressed with verbal protocols, both examining existing hypotheses and generating data on more general issues.

5.1 INTRODUCTION

The processes underlying response to attitude questions are not well understood. In part, because there is a paucity of actual process data. This study aims to address this lack of understanding by using verbal protocols to generate data on response processes with attitude questions. There are a variety of factors which may affect the processing of attitude questions, including the particular wording of a question, the type of response scales provided, the order of questions, and the context in which the question is asked.

Much of the research on attitude questions within the CASM framework has focused on context effects. However, rather than concentrate on this one factor, context, this study aims to explore a number of factors which may influence responses to attitude questions. The aim of this approach is to provide information on a range of factors to which think-aloud may be

applicable, and also to explore the processing involved in responding to attitude questions over a range of issues.

As each set of questions used in this study deals with a different issue, each will be treated as a separate experiment. Each issue will contain an introduction to the problems, the results of the experiment, and a discussion of the results. The method, procedure and respondents used is the same as for the previous chapter (see pages 157 - 159).

The first questions to be dealt with here concern life satisfaction and deal with a clear hypothesis about the nature of assimilation and contrast effects. The questions are drawn from previous research which has shown response effects for these questions on split-ballot surveys. A specific hypothesis developed from split ballot surveys was tested. This will be described below.

The other questions used were not drawn directly from work done with large scale split-ballot surveys. Rather, drawing on theories of responding, a number of areas where response effects might be expected are examined. These include the following questions: satisfaction with democracy, which varies the scale used to look at the issues of mapping responses onto a response scale; questions concerning European Union membership, which looks at the issue of question order; and a question on post-materialist values, which looks at the measurement of values and their stability between different response alternatives which, supposedly, measure the same

thing. Finally a general discussion of response processes at attitude questions, based on the findings from the above questions will be presented.

5.2 ASSIMILATION / CONTRAST EFFECTS FOR LIFE SATISFACTION.

5.2.1 Introduction

Previous research has found that asking a question about marital happiness before a question about general happiness led to respondents describing themselves as less happy generally than when the general question came first (Schuman & Presser, 1981). It was suggested that when the marital question is asked first, marital happiness is subtracted from general happiness. However, others found the reverse (Smith, reported by Tourangeau, Rasinski, & Bradburn, 1991). In this case the marital question is seen as priming marital happiness, making it accessible. Strack, Martin, and Schwarz (1988) hypothesised that these effects may be due to the use of conversational norms. Depending on the conversational context of the questioning, this context may cause the specific area to be seen as part of general life satisfaction, and therefore relevant to the evaluation of general life satisfaction. As the area has just been mentioned it is accessible and therefore more likely to influence the judgment of general satisfaction, producing an assimilation effect. On the other hand the conversational context may cause the general question to be seen as asking for new information,

apart from the specific area already reported on. Here the specific area is discounted from the judgment of life satisfaction, producing a contrast effect.

Schwarz, Strack, and Mai (1991; also Strack et al 1988) demonstrated this by asking about dating satisfaction before general life satisfaction either with a lead-in to the questions which suggested the questions represented two separate domains or without this lead in. When the questions were asked without a lead in there was greater correlation (.67) between the two areas of satisfaction than when a lead in was used (.18). Tourangeau, Rasinski, and Bradburn (1991) provide similar results when the specific area is marital happiness.

If this hypothesis is correct then in think aloud protocols those respondents who do not receive a lead in should refer more to evaluations of their relationship when evaluating overall life satisfaction than when there is a lead in. In addition, discounting of the relationship satisfaction should occur when a lead in is present but not when it is absent.

In this experiment the questions and lead in from Strack et al were used, but with the more normal 3-point verbal scale (rather than 11 points as used by Strack et al; Tourangeau, Rasinski, & Bradburn, 1991). These are given below.

Figure 5.1
Life satisfaction Questions and Lead-in

```
LEAD- IN.
 I would now like to ask you to report on two aspects
 of your life, which may be relevant to people's
 overall well being:
     1) marital/relationship/dating satisfaction.
     2) satisfaction with life as a whole.

QUESTIONS.
1. Taking things altogether how would you describe
    your (relationship). Would you say that your are
    very happy, fairly happy, or not too happy with
    your (relationship)?

2. Taken altogether, how would you say things are
    these days? Would you say that you are very happy,
    fairly happy, or not too happy?
```

Note: a previous question was asked to establish marital
status.

5.2.2 Results

Responses

Relationship satisfaction and life satisfaction are positively
correlated for both groups. For the group with a lead in the
correlation is .52 (p=.032); for the group without a lead in
the correlation is .75 (p=.000). Thus, the difference in the
size of the correlations between the two groups is in the
direction expected, but is much less than that found by
Schwarz et al.

Protocols

Analysis of the protocols showed very few explicit inclusions
of relationship at the general happiness question; 3 people
included their relationship in the group without a lead in and
2 in the group with a lead in. There were no explicit
exclusions of relationship. A few respondents raised

questions as to whether the relationship should be included
but did not resolve these questions.

For the behavioural frequency questions, when faced with a
lack of data which directly addressed the hypothesis, we
turned to look at general response strategies. However, in
the case of life satisfaction, respondents, firstly, do not
say very much. And, secondly, what they do report is mostly
a general evaluation or affective response, sometimes with the
addition of one particular aspect (eg financial
circumstances). Only a few respondents break their
evaluations down into a number of areas of satisfaction or use
a comparative strategy (eg life now compared to life before).

Table 5.1
Strategies for life satisfaction

	NO LEAD-IN	LEAD IN	TOTAL
GENERAL	10	8	18
GENERAL + ASPECT	3	4	7
ASPECTS	4	5	9
OTHER	3	2	5
RESPONSE ONLY	1	1	2
TOTAL	21	20	41

5.2.3 Discussion

Although the finding that respondents' use of 'general
affective' statements in response to life satisfaction
questions is interesting in regard to satisfaction in general,

it is of little interest in regard to testing the present hypothesis.

Given the survey results from other researchers, one would expect that assimilation and contrast effects would have occurred in this study. If assimilation and contrast did occur here, then they did so at a non-reportable level. And, moreover, both inclusion or exclusion of relationship satisfaction have gone into the overall evaluation of satisfaction without, in most cases, the respondent's conscious awareness of this inclusion or exclusion. How the general evaluation is produced is not evident, and the influence of the previous relationship question on this evaluation is also unclear.

For the behavioural questions reported in the previous chapter, while the direct link between independent variable and response was not articulated by the respondents, there was a clear way in which the strategies used could be influenced by these variables. Here, however, neither the independent variable nor, for the most part, the 'strategies' seem reportable. Thus, the influence of conversational norms and of the accessibility of information, remains obscure.

5.3 COMPARING SCALES -- SATISFACTION WITH DEMOCRACY.

5.3.1 Introduction

Response scales are commonly used for attitude questions. Respondents are asked to rate, for example, not simply whether they agree or disagree with an item but the extent of that agreement or disagreement. Issues surrounding the use of scales concern whether they should be bi-polar or unipolar, how many points a scale should have, and the extent of verbal labelling of a scale.

Whilst it has been common to assume that people have dimensional representations of issues, Ostrom (1987) argues that people find a continuous rating scale difficult because they have categorical, all-or-none, representations. He argues that people decompose continuous scales into a categorical form. Thus for a bipolar scale they would first decide which side they are on, and then perhaps further decompose the scale. Ostrom, Betz and Skowronski (1992) report a study by Krosnick and Berent which they claim supports this idea. They report that Krosnick and Berent found that test-retest reliability was higher when respondents first selected the side they supported and then indicated how strongly they were on that side.

Within the CASM framework there has not been much work on the effects of different scales. 7 point scales (plus or minus 2) have generally been seen as the most reliable (Cox 1980).

Alwin (1992) reports that in general the more scale points the more reliable the scale, the exception being two category scales. Alwin and Krosnick (1989) also report higher reliability for scales which have more verbal labels. Krosnick (1991) suggests that verbal response alternatives are easier for respondents than numerical scales with only end-point labels as for the latter the meaning of the points in the middle of the scale are ambiguous.

To investigate cognitive processes and response scales a question was chosen from the Eurobarometer survey which has been asked with different response scales. The question asks people how satisfied they are with democracy in their country. The response scales given are either a 4 point verbal scale or a 10-point numeric scale with only the ends labelled. The question and scales are given below.

Figure 5.2
Satisfaction with Democracy Question

QUESTION WITH VERBAL SCALE
Are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the way democracy works in (your country)?

QUESTION WITH NUMERIC SCALE
On the whole, to what extent would you say you are satisfied with the way democracy works in (your country)?

Please use the scale on this card to indicate your reply. 10 means you are completely satisfied and 1 means you are completely dissatisfied.

completely   completely
dissatisfied   satisfied
   1   2   3   4   5   6   7   8   9   10

5.3.2 Results

Responses

Mean satisfaction with democracy for the verbal scale was 2.6 (SD =.88). Mean satisfaction for the numeric scale was 5.2 (SD=2.5). Both means are in the middle of the scales, and thus show little difference across scales.   2.5 & 5.5

Protocols

Many respondents thought of aspects of democracy or particular issues -- proportional representation, informational or cultural barriers to access, referenda etc. For example one respondent says " I was thinking mostly about the first past the post. The fact that you can rule with 42% of the vote." Another says "Student rights are basically non-existent, women's rights are non-existent. Racial equality is pathetic". Others used a comparative strategy, comparing democracy in Britain with democracy elsewhere or with other types of systems, for example "compared to other countries, the democracy in this country is very good". Others mention both particular aspects and comparisons. The table below gives a breakdown of these categories across groups.

Table 5.2
Strategies for satisfaction with democracy by group

	Verbal Scale	Numeric Scale	Total
Aspects/Issues	12	6	18
Aspects/comparisons	4	2	6
Comparative	1	2	3
General evaluation	1	5	6
Scale	1	2	3
Other	2	3	5
	21	20	41

Some simply give general evaluations of democracy, sometimes linked with an attitude towards democracy and sometimes with more concentration on the response scale. An additional category reflects those who concentrate simply on the scale rather than on the topic of democracy. An example of a general evaluation is: "You gotta be optimistic about democracy, its not a good thing but I'm going to give a good mark cause we have democracy and it's not good to spit on it so easy". And, one which focuses exclusively on the scale: "its something that's difficult to sort of put a numerical value on, but I suppose it's sort of somewhere between fairly satisfied and completely satisfied".

As can be seen from the table there are some differences between the groups. Those given the verbal scale bring in aspects or issues related to democracy more than do those who receive the numeric scale. These latter tend to have more varied strategies, and although some look at aspects, there is also more concentration on general evaluations and scaling.

These codes were collapsed in terms of responses which give some thought to issues about democracy versus those that are more superficial. Respondents who either thought about aspects of democracy or who used comparative strategies (the categories 'Aspects/issues', 'aspects/comparisons' and 'comparative' in table 5.2) were grouped into the 'considered' category. Those who gave general evaluations, concentrated on the scale or used other strategies were coded as 'superficial'. Those given the verbal scale are more likely to consider a particular issue, whereas those given the numeric scale are equally likely to consider an issue or to treat the question more superficially. The table below illustrates this breakdown.

Table 5.3
Strategies for satisfaction with democracy by group,
divided into considered versus superficial

	Verbal	Numeric	Total
considered	17	10	27
superficial	4	10	14
Total	21	20	41

5.3.3 Discussion of scaling

It seems that those who receive the verbal scale are more likely to think about the issues involved than are those who receive the numeric scale. Having said this the level of consideration is again not large. Most people mention only one issue, such as proportional representation, rather than considering a number of relevant aspects. A few do mention

both negative and positive aspects; and a few also bring in comparative arguments.

Why do we get this difference between the groups in terms of mentioning issues? Firstly, it should be noted that there are a number of differences between these scales. One is 4 point, one is 10 point; one is numerical, one is verbal; one is presented orally, one on a showcard. Any one of these, or combination of them, could be responsible for the differences. It is also possible that for the numerical scale respondents thought about aspects whilst the question was being read but then were distracted by the show card and forgot this thinking. Thus this result might be an artefact of the think aloud procedure, however, the retrospective report may have allowed them to recall this thinking. Another plausible reason is that the numerical scale is indeed distracting, but it distracts the respondent from thinking about the question in terms of the issues of democracy and focuses them onto the issues of placing themselves on a scale. The verbal scale is more comprehensible, it is closer to ordinary language and thus does not represent such a difficult response problem. Respondents using the numeric scale use ordinary language terms like 'fairly satisfied' before translating this into numbers, suggesting an additional response step is required. This idea that it is the numeric aspect of the scale which is more difficult is consistent with Krosnick's (1991) suggestion. He assumes satisficing will result from more difficult scales. Perhaps part of this satisficing is not simply in choosing the response, but in putting effort into

thinking about the issues. Further research would be needed to determine which particular aspects of these scales make them more or less difficult.

A further issue in terms of the numeric scale is whether people tend to decompose the scale into categorical form as Ostrom (1987) suggests or whether they deal with it more as a continuous scale. For some respondents one cannot really tell how they have dealt with the scale; there is no trace of this process in their protocol. About half the respondents do however, show some type of scaling statements. A few respondents do seem to immediately place themselves on one side of the scale; for example, "I would probably go on the side of slightly dissatisfied" or "I was thinking I'll choose a higher number, it had to be higher than 5 or 6 because it's more important to me than something mediocre or average". Also notable in the above quotes is that people use verbal labels or categories to roughly place themselves things like "pretty dissatisfied" or "fairly satisfied" occur for a number of respondents. They then, presumably, have to choose a number which seems to fit that description. But quite how the verbal judgement is mapped onto numbers is usually elusive. A couple of respondents indicate choosing because it sounds right " I chose 7 because its a nice in between number". A few respondents do seem to provide at least one endpoint exemplar, for example, " we're not a dictatorship so you can't be completely dissatisfied". A couple, though, seem to use both sides of the scale to try to find their position, though not

explicitly so; for example, "feeling that its neither not not the worst but not the best".

Thus whilst for about half the protocols there is no trace of scaling their responses, for the others there is evidence of the way they approached the scale. More seem to place themselves on one side of the scale than to consider both ends, this would support Ostrom's idea that people approach the scale in a more categorical way. Additionally, the use of verbal labels can act as further subdivisions of the scale. What may differentiate people who simply consider one side of the scale from those who consider satisfied and not satisfied dimensions may be the strength, or extremity, of their attitude to democracy. Whilst the protocols don't really provide evidence for this, it could be that those who know their general position can immediately go to one side of the scale or the other, whilst those who have not really considered their position before need first to consider which side they might be on.

5.4 QUESTION ORDER -- COGNITIVE AND AFFECTIVE RESPONSES TO EUROPEAN MEMBERSHIP.

5.4.1 Introduction

The order in which questions are asked can effect responses to questions. Such order effects are generally termed context effects. Much of the work in CASM on context effects and attitude questions focuses on assimilation and contrast

effects for two types of context: part-whole question sequences, of which the life satisfaction questions above are an example, and related area priming, where a related area which is either positive or negative to the target issue is first presented and acts to prime different sides of an issue.

Context may also be important where different aspects of the same issue are queried. It is often argued that one question is unreliable for measuring an attitude and that many different questions assessing the same attitude are necessary. The three component model of attitudes suggests that attitudes are a combination of affective, cognitive and behavioural components (Rosenberg & Hovland, 1960). When measuring attitudes one needs questions assessing these different aspects, and often surveys attempt to do just that.

But, might measuring one aspect affect responses to the other aspect? This may be the case, especially if the different components are inconsistent, which has been shown to make the attitude less stable (Rosenberg, 1960). Asking people to think about why they feel the way they do has been shown to lead to attitude change (Wilson, Lisle, & Kraft 1990; Wilson, Kraft, & Dunn 1989). Thinking which focuses on one aspect of the attitude (affective or cognitive) can lead to less relation between attitude and behaviour when consistency between cognitive and affective aspects of the attitude are low (Millar & Tesser 1989, 1992). Similarly, informational messages have been shown to lead to more change in affective attitudes and emotional messages to more change in cognitive

attitudes (Millar & Millar, 1990). It may be that simply asking a question which makes one aspect of the attitude more salient or accessible, will affect responses to subsequent questions assessing a different aspect. If first questioned on their beliefs people may use the accessible information, about beliefs, when answering the later affective question, and vice versa. Thus one might expect that the order of presentation of items tapping different aspects of the same issue may lead respondents to answer differently. This might be expected to produce a response effect when affective and cognitive aspects are divergent.

To look at the effect on processing of the order of questions dealing with affective and cognitive components of attitudes two questions related to European Union membership were asked. Respondents were asked for their reaction to the EC being scrapped and whether the EC had benefited Britain. SCRAPPED is meant to tap the affective component of attitudes towards Europe while BENEFIT is meant to tap cognitive components. Both of these questions are taken from the Eurobarometer survey, and are given below.

Figure 5.3
Questions on attitudes to Europe

| Q. If you were to be told tomorrow that the European community had been scrapped would you be very sorry about it, indifferent, or relieved? (SCRAPPED) |
| Q. Taking everything into consideration, would you say that Britain has on balance benefited or not from being a member of the European Community? (BENEFIT) |

In this experiment the order in which the questions were asked was varied. Half of the respondents received the SCRAPPED question first, while the other half received the BENEFIT question first.

5.4.2   Results

Responses

The tables below give responses to BENEFIT (table 5.4) and SCRAPPED (table 5.5). For both questions there is little difference between the two groups.

Table 5.4
Responses to BENEFIT by group

	Benefit	Not Benefit	DK	Total
BENEFIT first	13	4	3	20
BENEFIT second	17	2	2	21
TOTAL	30	6	5	41

Table 5.5
Responses to SCRAPPED by group

	Very Sorry	Indifferent	Relieved	DK	Total
SCRAPPED 1st	14	5	2		21
SCRAPPED 2nd	10	7	1	2	20
TOTAL	24	12	3	2	41

Protocols

For SCRAPPED the two largest categories are evaluations of the EC or affective reactions, and considering the effects of

scrapping, either personal or social effects. The table below (5.6) shows that there are some differences between the group who were asked this question first and those who received it after 'benefit'

Table 5.6
Strategies for SCRAPPED by group

	SCRAPPED first	SCRAPPED second
Evaluations / Affective reactions	9	1
Effects on personal/ societal life	6	13
Evaluation plus specific aspects	2	1
No knowledge / interest in area	3	2
Other	1	3

Those respondents who answer SCRAPPED first tend to talk in terms of affective reactions, or general evaluations of the EC and simply whether the prospect is good or bad. Those who answer SCRAPPED after BENEFIT talk more in terms of the effect of scrapping either on their personal life, or everyday life in general or in terms of the effect for the country.

For BENEFIT, many respondents refer to some aspect in which Britain has or has not benefited. These include economic, political and social /cultural areas. For example, "I would say that the way the European community's structured that it will benefit the it's designed to benefit larger economies like Britain" or " I was thinking about the education system

in Britain, how they may have benefited in coming getting closer to the continental way of educating and how they'll probably also benefit from knowing other languages".

Of these respondents who consider aspects of BENEFIT, most consider only one issue while about a third consider more than one issue. The aspect most mentioned is economics. Economic aspects are generally seen more as beneficial than not. Of those few who consider more than one aspect most consider at least one negative and one positive aspect.

Social aspects are seen as beneficial by all who consider them. The only slight difference between the groups who receive BENEFIT first and those who receive it after SCRAPPED comes from the mention of social aspects. Seven people mention social aspects; six of these are from those who received BENEFIT after SCRAPPED with only 1 when BENEFIT comes first.

Respondents who do not consider particular aspects can largely be said to answer superficially, with a number of types of general statements, a few attitudes to the EC and assertions of complete lack of knowledge on the issue. The general statements range from simple assertions, saying Britain would be worse off it they weren't a member as well as more idiosyncratic responses. For example, "I can't think of anything specific, well I don't, I think just on the whole they have"; or " Oh actually that's very difficult for me to judge. I would just say that everybody benefits from it so".

Many respondents express a lack of knowledge on this issue or uncertainty about their response. Almost all those who express a lack of knowledge do not consider aspects of benefit. A few respondents who admit to a lack of knowledge give don't know responses, most, though, try to generate some type of response. Table 5.7 below gives a breakdown of these codes across groups.

TABLE 5.7
Strategies for BENEFIT by group

	BENEFIT first	BENEFIT second
Aspects of benefit	10	8
General statements	7	9
Attitude to EC	1	2
No knowledge	1	2

In summary then, a number of people consider specific aspects of benefit, but few consider many aspects. The only minor difference between the groups comes from a greater mention of social aspects from those who receive SCRAPPED first. Many people lack specific knowledge or give vague responses, but are still prepared to give an opinion. This certainly goes against any model that sees people as weighing up the pros and cons of an issue and forming a judgement on this basis.

5.4.3 Discussion

The small differences between the groups suggests that asking about affective reactions first has little implication for the

question of BENEFIT -- the more cognitive appraisal of the issue. The slight difference in more mentions of social aspects when SCRAPPED is asked first may come from the more positive nature in which these issues are seen. People see being involved with other countries as positive. Perhaps having first brought to mind one's affective reaction makes these considerations more salient, and more likely to be considered as benefits. In this case then the response process does not seem to be affected, but different contents may be made more salient.

For the question SCRAPPED though it seems that it does matter whether the question is asked first or second. Perhaps bringing to mind an affective response to an issue, or at least to this issue, has less influence on cognitions about it than the reverse order. There is some suggestion that where an attitude has a large affective component, concentrating on cognitive aspects may reduce the attitude-behaviour consistency (Millar & Tesser, 1992). Certainly here having considered more cognitive aspects of European membership seems to affect thinking about the affective reaction question. It is also true that few people seem to have much specific knowledge of the EC, even though they have an opinion about it. This issue may be dominated by affective reactions rather than by cognitions. In this case focusing on the cognitive aspects of the attitude first may lead people to think about the affective aspects differently than they otherwise would have.

From the small sample there is no evidence for this difference in thinking translating into a response effect. But, this study is consistent with studies which show effects on attitude-behaviour consistency or attitude change depending on whether people focus on cognitive or affective components of an attitude (Millar & Tesser, 1992; Millar & Millar, 1990). The present study shows how the different focus can lead to different ways of thinking about the issue. For the questionnaire designer this is problemmatic, because a simple change in the order of questions could potentially produce differences in response. A response effect is probably most likley where the different components are inconsistent because in this case the different components have different implications for response.

5.5 MEASURING VALUES -- RESPONSE OPTIONS FOR POST-MATERIALISM.

5.5.1 Introduction

Values are assumed to be more basic and enduring than attitudes. Indeed attitudes are sometimes seen as deriving from values (Rokeach 1973). But, to determine what values people have, one still has to measure them in some way. Usually values are measured by asking respondents to rank issues, for example from most to least important. Rankings are assumed to reflect the underlying value structure, that is, those things ranked highly reflect what is valued. Krosnick and Alwin (1988; Alwin & Krosnick, 1985) suggest this

may be better than rating each issue separately. Ranking, being a more complex task, may demand more thoughtful responses from respondents. In the present study a particular type of value is examined, Inglehart's materialist / postmaterialist value dimension.

Inglehart (1977, 1990) proposed a shift in values to account for what political scientists see as the emergence of a new political agenda in western democracies. This new political agenda involves the concentration on a new and expanded range of issues in politics such as women's rights and the protection of the environment. Inglehart saw this value shift as a movement away from materialist values towards post-materialist values; from an emphasis on economic stability and security, to values which emphasise the need for participation in society, human rights and aesthetic needs. Whilst his thesis has changed in terms of the origins of this value change, at first linking these needs to Maslow's (1954) hierarchy, and later in terms of diminishing marginal utility, the basic theme of a change from materialist to post-materialist values has been maintained.

The instrument for measuring post-materialist / materialist values consists of asking respondents to rank what they consider should be the long term priorities for their government. There are two versions of the scale -- the short version and the long version. The short version consists of one four item list with two materialist and two post-materialist values. The long version consists of three such

lists. The short version uses the list labelled A below (see figure 5.4), whilst the long version consists of this list as well as two others, including the one labelled B below.

FIGURE 5.4
Question and response options for
Materialist / Post-materialist Values

Q. There is a lot of talk these days about what a country's
   goals should be for the next 10 or 15 years. On this card
   are listed some of the goals that different people say should
   be given top priority. Would you please say which one you
   yourself consider most important for your country in the long
   run.   And what would be your second choice.

       LIST A
   Maintaining order in the nation
   Giving the people more say in important government decisions
   Fighting rising prices
   Protecting freedom of speech

       LIST B
   Maintaining a high level of economic growth
   Making sure that this country has strong defence forces
   Seeing that people have more say about how things are done at
    their jobs and in their communities
   Trying to make our cities and countryside more beautiful

The materialist / post-materialist items are asked within a survey context, notably the Eurobarometer survey has included the questions for a number of years. Respondents are labelled 'post-materialist' if they choose the two post-materialist options, 'materialist' if they choose the two materialist options, and mixed if they choose a mixture.

This does seem to be a rather simple, one might say simplistic, measure of values. Yet the fact remains that both across time and across cultures, many of the expected relationships with this measure are supported, for example a relationship with social background, and with opinions on a variety of issues, including support for women's rights

(Inglehart, 1981). Indeed, there is an impressive amount of data which support this measure of values.

Inglehart's thesis has been criticised on a number of grounds, both theoretical and methodological. It is the latter which concerns us here. Flanagan (1982; Inglehart & Flanagan, 1987) claims that it confounds two dimensions: a materialist / post-materialist dimension and an authoritarian/liberal dimension. And Van Deth (1983), using panel study data, showed that whilst the measure may be stable at the aggregate level (proportions in different age cohorts as predicted) there is little stability at the individual level. Interestingly, those with less interest in politics showed less stability. This lack of individual stability has been put down to measurement error, however this is to dismiss the problem too easily. In this case the presence of measurement error raises serious questions about what exactly is being measured by this value scale.

In the present study this issue of what is being measured by the materialist / post-materialist scale was explored with think aloud protocols. To do this respondents were presented with different lists of response options. One group received the list labelled A above; the other group received list B. The different lists are meant to measure the same basic values, and the selection of an item is presumed to be guided by one's underlying values. Thus, if a postmaterialist value orientation is guiding response selection, the particular wordings of different postmaterialist and matierialist options

should not affect response. The options should be evaluated in terms of of the value orientation. At the value level the different respective options are equivalent. This is unlike the proposition for attitudes that different wordings may be tapping different aspects of a multifaceted issue and hence may produce different responses. What is being proposed for these values is that different wordings are inconsequential.

5.5.2 Results

Responses

First, looking at the actual responses given by respondents to the question (that is whether they chose materialist or post-materialist options) there were differences in responses between the two groups. As can be seen from table 5.8 below, respondents who were given list A were significantly more likely to be classified as post-materialist than those given list B (X^2 = 12.07, df=1, p=.0005). Note that the number of materialists in our sample was too small to analyse statistically and therefore 'materialists' in the table below includes respondents who chose at least one materialist item.

Table 5.8
Classification of respondents as materialist or
postmaterialist by group

	LIST A	LIST B	TOTAL
MATERIALIST	3	14	17
POST-MATERIALIST	18	5	23
TOTAL	21	19	40

Protocols

Verbal protocols resulting from think aloud were coded in terms of how people dealt with each item in the list. Categories included whether people considered aspects of each option, whether they gave simple evaluations or simply labelled the option, and whether they gave no consideration to the option (they did not mention it or merely read it out without further comment). Table 5.9 below shows that the respondents presented with list B gave more consideration to aspects of each option than did those who responded to list A, who tended not to evaluate as many options.

Table 5.9
Protocol Codes for List Options of Materialist /
Postmaterialist values by group

	LIST A	LIST B	TOTAL
Aspects Considered	20	34	54
Simple evaluation/			
Labels	34	27	61
No consideration	29	18	47
other	1	1	2
TOTAL	84	80	164

The protocols suggest that respondents given list A tended to think less about the options than did respondents given list B. Whilst this relation obtains over all the options, the largest differences along these lines comes from the evaluation of the economic options ('rising prices' and 'economic growth'), with 'rising prices' being quickly eliminated but 'economic growth' receiving much consideration. The security options ('law and order' and 'defence forces') on

the other hand, tend to be more similarly evaluated on both lists.

This idea that list A respondents are thinking less than list B respondents is further supported by an analysis of the amount of verbalisation. One list A respondent talks a good deal more than other respondents. If this one extreme respondent is eliminated then there is a significant difference in the amount of verbalisation produced between the groups, with list B respondents verbalising more than list A respondents (t=2.56, df=38, p=.015).

Also, in list B, the post-materialist option 'beauty' is less likely to be chosen. It is, for many respondents, seen as rather trivial and sometimes as dependent on economic growth and other options. In this interpretation the 'beauty' option is seen as being about nice architecture and cleanliness. One respondent for example says "I think people can find beauty in the cities and countryside if they want to. It's a question of outlook I think. Maybe if they've got more confidence in their social worlds and like the personal world as well, which comes from that, then they'll be able to see beauty in more things." However, those few who chose this option interpret it differently. They see it as something which 'could be' about protecting the environment. Hence, the meaning of this item is ambiguous: does it refer to keeping the cities sparkling clean or preserving the natural environment?

Furthermore, the economic growth option is often evaluated in an instrumental way; it serves to make other, valued, things possible. For example, one respondent says "economic growth is a means to an end and having everyone happy is basically what you're trying to do". Also, some respondents answer not simply in terms of what they want, but what is realistic. For example one respondent says "I wanted to go for 3 [say in jobs] cos it actually seemed nicer, but it wasn't in a sense as realistic as maintaining a high growth. ......In a hard way it seemed to be 1 [economic growth] cos it seemed to me one that any others might be founded on."

A further look at post-materialism: The dimension underlying the measurement.

To supplement the think-aloud analysis of this question we also explored the dimensions underlying the question items. Central to the measurement of materialism/post-materialism is the idea that there is a dimension underlying the construct. Using Inglehart's technique for measuring this presupposes that agreement or disagreement on all the options in lists A and B fall along the same continuum. In order to investigate if this is true we had each respondent, after taking part in the think-aloud, make semantic differential ratings between each pair of options. For example, each respondent was asked to rate, on a scale from 1 (very similar) to 9 (very dissimilar), how alike 'Making sure that this country has strong defence forces' and 'Maintaining order in the nation' were. We split the 'say in jobs and community option into two ('allowing people more say about how things are done at their

jobs' and 'allowing people more say about how things are done in their community') because we felt these might be tapping distinct beliefs. Thus, there were nine options and therefore thirty-six pairs to be evaluated. The pairs were presented in the optimal order for paired comparisons (see Wells, 1991).

Two of the original respondents were unable to complete the task and one respondent's ratings were so disparate that it is likely he or she failed to understand the procedure. The analysis is of the remaining 38 respondents.

Multi-Dimensional Scaling (MDS) is a technique well suited for exploring such a data set (see Jones & Koehly, 1993). Essentially this technique attempts to map out any underlying dimensions on which the sematic differential judgements were made. Because we had 38 respondents we used a type of MDS known as INdividual Difference SCALing (INDSCAL). This allowed us to compare individual respondents. With the exception of the single respondent who failed to understand the task, there were no outliers or systematic variations. Thus, only the results for the group data will be presented.

The main output from MDS is a graph of the items giving their positions in what is argued to be the space for how the semantic ratings were made. In almost all cases MDS is used as an exploratory technique. A two dimensional solution fits this data adequately (see figure 5.5. The stress value is .426. The overall importance of both dimensions is similar:

dimension 1 has an importance value of .234 and dimension 2 a value of .207).

The first dimension does distinguish the materialism / post-materialism options well. However, there appears also to be another dimension which separates the 'Beauty', 'rising prices', and 'economic growth' options from the others. This makes sense if we take the 'beauty' option to be keeping the cities and countryside clean and prosperous. As a number of respondents in the think aloud part of the study suggest, respondents may see a clean and aesthetically appealing environment as linked to economic prosperity; beautiful cities may be seen as an aspect of wealth. Prosperous places are beautiful places.

Figure 5.5
Examining Value Options.
Derived Stimulus Configuration for Two
Dimensional Solution.

Dimension 1 (horizontal)   vs   Dimension 2 (vertical)

```
        |  |   |   |   |       |   |   |   |   ||
        |                     Prices            |
        |      Beauty 9        35 Econ          |
  1.0  -+                                      +-
        |                                       |
        |                                       |
  0.0  -+---------------------+-----------------+
        |   Saycom 8          |                 |
        |        7 Sayjobs    |     1 Order     |
        |                     |                 |
        |   Saygov   2        |     6 Defence   |
 -1.0  -+   Speech   4        |                +-
        |                     |                 |
        |  |   |   |   |      |   |   |   |   ||
       -2.5    -1.5    -0.5    0.5    1.5    2.5
```

- 225 -
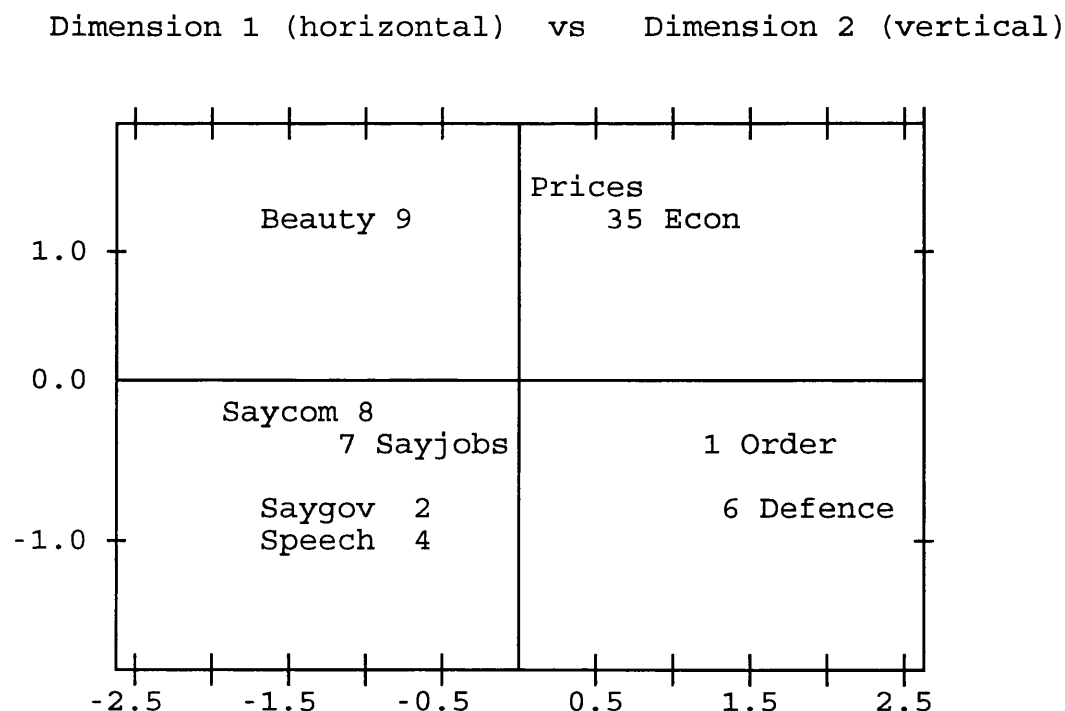
5.5.3 Discussion of postmaterialist values

The responses to this question show that the two lists are not equivalent. Different numbers of materialists and postmaterialists are produced from the different lists. In part this may be due to the ambiguity of the 'beauty' option. This option is problematic in that it is open to different interpretations, interpretations which mean that it is not necessarily seen as a postmaterialist value. This is apparent from both the semantic differential ratings and in considerations of the option during response. Whilst the semantic differential ratings do offer some support for Inglehart's thesis, they also show this problem with the 'beauty' option.

But ambiguity with the 'beauty' option does not offer a complete explanation of why the economic option, 'economic growth', is then chosen. The 'economic growth' option is seen as desirable by those who choose it; it is not necessarily simply the best of bad options. In fact many respondents choose it as their first choice. Many respondents see 'economic growth' as necessary for other things, including the other items on the list. For example one respondent says it "allows everything to happen" or "without a high level of economic growth you can't have people choosing which jobs they're going to do". Thus, for many, it is an instrumental value, a means to an end.

The option 'economic growth' then is positively evaluated by many, unlike the economic option on the other list, 'rising prices' which is quickly dismissed. Thus these two economic options in particular, but more generally the lists as a whole, differ. Part of this difference is the amount of consideration given to the options, as revealed by the differences between the groups both in the protocol codes and the amount of verbalisation. List A respondents seem to respond more 'automatically' while list B respondents are more thoughtful. What is it about the lists which influences the amount of thought? It is perhaps the case that the options on list A form clichés, or slogans, which are quickly interpretable by respondents, while options on list B do not have the same instantly recognisable quality and therefore require more consideration. Some list A respondents mention their quick assimilation of options; for example, one respondent says "two of them were conservative ideas, two of them were democratic ideas" and goes on to say about the conservative ones "I didn't think about them, I knew they were steeped in dogma". Others are similarly able to place labels on the list A options, much more so than those given the list B options. Thus there appears to be a difference in the recognizability of options.

5.6 GENERAL DISCUSSION

For most questions differences in the strategies used were found between the different groups and the implications of this have been discussed. What is not clear, in the case of

satisfaction with democracy and the European questions is whether these differences in strategy translate into response differences. Clearly one might arrive at the same response through different strategies. Testing in a survey context with larger samples would be needed to examine this issue.

For the one question, life satisfaction, where response differences have been found in large scale surveys, and where a clear hypothesis regarding the reasons for these differences exists, we found neither any direct evidence for the hypothesis nor any strategy differences. This does not suggest a failure in the hypothesis, but rather points to the limitations of verbal protocols. Given that differnet response profiles were found in survey research it must be assumed that the processes are not conscious. Where the underlying processes are not conscious and furthermore where nonconscious processes fail to influence the conscious strategy adopted, protocols fail to provide useful data for hypothesis testing.

Looking more generally at the thinking behind attitude questions, one theme running through the literature is that people sample from their beliefs about issues in order to respond. In the protocols we do indeed find evidence that people think in this way; many do consider aspects of issues. They think about proportional representation or treatment of minorities when thinking about democracy. They think of economic statistics or language use when considering the benefits of the EC. They give examples of the lack of freedom

of speech or the benefits of economic growth. Yet, even for those who do draw on their beliefs in considering aspects of the issue in question, the number of beliefs recruited is, generally, not large; often only two or three, sometimes only one aspect, sometimes more.

This limited sampling may be cause for concern. Central is the issue of salience; that is, the beliefs drawn on are those which are salient. One might expect that with such limited belief sampling anything that makes a belief more momentarily salient would be likely to influence response because few other beliefs are likely to be drawn upon. But what makes a belief, or aspect, salient? Some may be chronically salient, strongly linked to the issue; these would presumably be less swayed by momentary influences. Salience is, in part, a function of the social and personal salience of beliefs, instances, aspects etc. Certainly in the protocols people mention the media, both in general and in specific instances of articles they've read. They also mention conversations and personal experiences -- "I was talking about this recently". These aspects of salience are beyond the control of the questioner, although, there is scope to examine the media coverage of topics which one might be interested in.

Aspects which are made salient by previous questioning is more concerning for the questionnaire designer, in that beliefs made salient by previous questions may unduly influence later questions.   Especially,   as   noted,   because   so   few considerations are brought to bear on a question; people do

seem to answer with whatever information is readily available. This influence of previous questions may for instance occur somewhat for BENEFIT and SCRAPPED. When BENEFIT is asked first people seem more likely to frame the affective reaction in terms of loss of benefits. And, to a lesser extent for BENEFIT, when asked after SCRAPPED, people are more likely to consider social/cultural aspects as benefits, otherwise benefits are seen mostly in economic terms. Having thought about their reaction in terms of personal loss -- travel abroad, liking links with other cultures -- at SCRAPPED people are more likely to consider these aspects as potential benefits.

There is also evidence about the varying nature of attitude strength. Some respondents report having thought about the issue before. A number do so for life satisfaction; it is something they consider about themselves in daily life. But at other questions also some respondents indicate having thought about the issue before. This sometimes produces less thinking "I didn't need to think I just knew because I've thought about this a lot" but it also produces more rehearsed arguments. This may be especially true for social science students (some who study the European Community for example). Whilst this sometimes leads to a definite response, in other cases it does not. One respondent, for example, indicates having thought about the issue of European benefits, and sees economic benefits, but also considers that this is in the long run detrimental to the environment -- he responds 'don't know' because he has not resolved this problem. Another, with

exactly the same conflict, decides to give greater weight to the environment. Both of these respondents clearly are knowledgeable, but mixed in their attitudes, a category which Tourangeau, Bradburn, Rasinski, and D'Andrade (1989a, 1989b) identified as being susceptible to response effects. Perhaps different contexts would push these respondents to resolve the conflict in different ways.

But, as well as the many respondents who do consider aspects of issues, there are also many who do not. Some of these may be people with strong opinions who simply do not need to think about the issues, but many more are vague in their response strategies. A question here is identified by the differences in amount of thought about the postmaterialist question, where one response list seems to generate more reactions and one to generate more thought. Are these people with vague response strategies simply going on a 'gut reaction': "I like Europe so I'll say benefit", "democracy's a good thing so I'll give it a high mark" . And this raises the question of what we are trying to measure when we measure attitudes, and perhaps values. Are we looking for a reaction, or are we looking for a more thoughtful considered response?

This difference between rapid and thoughtful response mirrors the split between measurement in the psychometric versus the survey tradition. In the former, respondents are urged to not think but to respond quickly, whereas in the latter respondents are often urged to think carefully. A danger in the former approach may be that respondents are simply

reacting to well worn phrases, or vague momentary impulses. A danger in the latter is that they are formulating responses to issues they have not previously considered and may therefore be subject to a host of situational variations. Clearly this issue warrants further research.

Across both the behavioural and attitudinal questions protocol analysis has provided useful information for examining response processes. The use of protocol analysis in an experimental design has shown differences in response strategies and in this way provided information on a number of hypotheses concerning response effects. However, limitations of the use of protocol anlysis were also encountered. Chapter 8 provides further discussion of how this study contributes to an evaluation of the strengths and weaknesses of protocol analysis as a method for investigating response processes.

CHAPTER 6 - EXPERIMENT 3 PART 1

CONTEXT EFFECTS AND ATTITUDE QUESTIONS

SUMMARY

This and the following chapter report on an experiment examining factors involved in context effects in attitude surveys. In this study a standard split ballot experiment was used, with two versions of an attitude questionnaire, varying the context in which target questions appear, presented via computer. Instructions on 'how' to think were also varied, and latencies collected as a measure of processing speed. Thus there were two dependent variables: the responses to the questions and the latency of response. This chapter describes the results for the response variable. The next chapter reports on the results for the latency variable.

6.1 INTRODUCTION

The effect of context on attitude questions is a central issue for survey methodology. Much work in CASM has focused on identifying factors which influence whether or not the context produces an effect. These factors have been discussed at length in chapter one. The present experiment aims to explore some of these factors, taking up some issues which have received a good deal of research attention and some which, although they have received theoretical attention, have had less empirical exploration.

A number of person factors have been shown to influence context effects. In particular the attitude structure, in terms of the importance of an issue to the person and whether they are conflicted on the issue, has been found to be influential (Tourangeau, Rasinski, Bradburn, & D'Andrade 1989a, 1989b). Although 'attitude strength' components have shown mixed results (Krosnick & Schuman, 1988), Tourangeau et al suggest that those who have important but conflicted attitudes are more susceptible to response effects.

Although it has been theorised that the amount of knowledge an individual has about an issue plays a role in context effects (Feldman & Lynch,1988), this factor has received less attention within research on context effects. Research from other areas might suggest that having more knowledge would lead to less effect. For example, greater knowledge has been associated with greater resistance to attitude change (Wood, 1982), and with greater attitude-behaviour consistency (Kallgren & Wood, 1986). But it is not clear whether this would also lead to less context effects. It could be argued that having alternative inputs available would mean less reliance on context information; however, more knowledgeable respondents may also be more likely to recognise relations between context and target items. The three personal factors of attitude importance, conflict, and amount of knowledge will be examined in this study.

It has also been assumed that the type of context in terms of how related it is to the target issue is also important in

influencing context effects (Feldman & Lynch 1988). Whilst
some research has used correlations of context and target as
evidence of relatedness (Tourangeau et al 1989a, 1989b), this
issue has not received much empirical attention. Tourangeau
et al use issues which are all moderately to highly correlated
with the target issue. While they investigate the question of
context effect when the issues are highly related, they do not
address differences in the degree of relation. This issue
will be addressed in the present study. A further factor
concerning the type of attitude is the degree of familiarity
of an issue. Theoretically one would expect more familiar
issues to be less susceptible to context effects than less
familiar issues, but again there is limited evidence for this
in the survey literature. Issue familiarity will also be
addressed in this study.

One of the more situational factors suggested as influential
to context effects in a number of ways is the amount of
cognitive effort devoted to response. Lack of cognitive
effort is suggested to lead to satisficing strategies
(Krosnick 1991). It is also associated with the direction of
the effect. Assimilation is assumed to take less effort than
contrast effects (Martin, 1986). Within the survey
literature, there have not been many studies which address
this issue (Martin & Harlow, 1992).

Cognitive effort is often linked to the amount of time spent
thinking about an issue (Krosnick, 1991; Martin & Harlow,
1992). More time spent thinking is associated with more

effort and little time spent thinking with less effort. Thinking time then is potentially an important factor in context effects. Clearly though, the use of a think aloud technique would be inappropriate when thinking time is central. Although thinking aloud can be assumed not to alter thinking substantially, it does tend to increase the amount of time taken to respond.

In this study thinking time is manipulated by providing instructions on 'how' to think, in an attempt to influence how much people think about an issue, and to determine whether thinking effort influences response effects. Thinking time, or response latency, will be measured. This provides a check on whether the manipulation has worked, but also, as discussed in the next chapter, it provides some trace of processing.

6.2. METHOD

6.2.1 Design
The present study was designed to investigate multiple factors which may contribute to context effects for attitude questions.

Context was manipulated across two questionnaires for twelve target issues. Two context items preceded each target item. Context items were designed to evoke different aspects of the issue which favoured one or other response pole. Half the respondents received two items evoking one side of an issue

and the other half received context evoking the other side. The full questionnaires are given in appendix 6.

The amount of time spent thinking about the questions was manipulated by instruction to the respondents. Half the respondents receiving questionnaire A were instructed to 'take their time; think carefully' while the other half were instructed to 'answer quickly; not spend too much time thinking'.

Thus, we have two manipulated independent variables in a 2 X 2 design: questionnaire context and thinking instruction.

Individual differences on knowledge and attitude factors were measured in a post-experimental questionnaire which asked respondents three questions about each target issue. These were:

1) How much they know about the issue   (5 point scale) [knowledge].

2) How important the issue is to them   (4 point scale) [importance].

3) Whether their views on the issue are one-sided or conflicted [conflict].

This questionnaire is given in appendix 7.

In addition to the 2 independent variables of context and instruction, within subjects the type of target question was varied in terms of the degree of assumed prior knowledge about or familiarity with the issue. Three knowledge types were

included: Unfamiliar issues, media issues, and familiar issues. There were 4 target questions for each knowledge type.

The type of context was also varied in terms of its degree of relation to the target question. It could be either obviously or subtly related. 6 target questions had obvious context and 6 had subtle context (2 obvious and 2 subtle for each knowledge type).

Piloting knowledge and context type
The factors of knowledge type and context type were derived from a pilot study of 40 students who completed a pen and paper questionnaire. Respondents were asked first how much they knew about each of the target issues. Knowledge was assessed by a five point scale from 'nothing at all' to 'a great deal'.

Second, each context question was paired with the relevant target issue and respondents were asked to rate how related the two issues were. They responded on a five point scale from 'not at all related' to 'extremely related'. This questionnaire is given in appendix 8.

Respondents were most knowledgeable about the familiar issues and, with one exception, least knowledgeable about the unfamiliar issues, with the media issues in between (see table 6.1 below). The one exception is the rather low score for knowledge about 'tradeoffs between job creation and damage to

the countryside' (labelled 'Jobs vs environment' below). It was felt however, that it was perhaps poor wording of the issue description that lowered the knowledge rating for this, rather than the issue per se.

Table 6.1
Mean knowledge of issues in pilot study

ISSUE	KNOWLEDGE TYPE	MEAN	SD
Food choices	Familiar	4.03	.698
Student housing	Familiar	3.80	.823
Course readings	Familiar	3.20	.823
London's cleanliness	Familiar	3.00	.889
EC benefits	Media	2.50	.847
Criminal justice	Media	2.45	.904
Pornography	Media	1.98	.920
Virtual shopping	Unfamiliar	1.93	.917
Jobs vs environment	Media	1.75	.899
Inuit whaling	Unfamiliar	1.65	.893
Opera funding	Unfamiliar	1.55	.904
Genetic engineering	Unfamiliar	1.50	.679

With the exception of 'course reading' the respondents rated the contexts which were meant to be obvious as more related to the target issues than the contexts which were meant to be subtle (see table 6.2 below). Because of the high relatedness ratings obtained for 'course reading' this set of questions was changed to make the context less obvious. Also, one side of the context for 'EC benefits' was rated as more related than the other; this side was changed to make it less obvious.

Table 6.2
Mean relatedness of context to target
(combined into issue area)

TARGET ISSUE	CONTEXT TYPE	MEAN	SD
Criminal justice	obvious	3.73	.79
Jobs vs environment	obvious	3.58	.715
London's cleanliness	obvious	3.46	.619
Course readings	subtle	3.62	.824
Genetic engineering	obvious	3.12	.745
Inuit whaling	obvious	3.01	.760
Student housing	obvious	2.97	.781
EC benefits	subtle	2.42	.646
Food choices	subtle	2.19	.945
Virtual shopping	subtle	2.81	.755
Pornography	subtle	1.98	.768
Opera funding	subtle	1.65	.585

Expectations for results

From previous research and theoretical discussions, a number of general predictions are possible of where context effects are likely to be found in this study.

In terms of the direction of the effect, those who are instructed to 'think carefully' should show more contrast effects, and those who are instructed to 'think quickly' should show more assimilation effects.

For the within subject factors of issue familiarity and context type the following pattern of results is expected: first, the more familiar the issue the less context effects are expected; second, obvious contexts should produce more context effects than subtle contexts.

For the individual factors, those who are both conflicted on an issue and who see the issue as important should show more context effects. Those with more knowledge of an issue should show more context effects for unfamiliar issues.

6.2.2 Procedure

100 students, recruited from around the LSE, responded to the questionnaire. Respondents were paid for taking part.

The experimenter gave a brief explanation of the task and then respondents were shown into a computer booth. Instructions for completing the questionnaire were given on the screen for respondents to read at their own pace. They were told that their opinions were requested on a number of issues. There then followed instructions on 'how' to respond which was one experimental variable. Half of the respondents received a set of instructions which encouraged them to 'take their time and think carefully about the issue before responding'. The other half received instructions to 'answer as quickly as possible'. The full instructions are given in appendix 9. Instructions on how to enter responses were also given.

Two practice questions were then given to familiarise respondents with the computer presentation before proceeding to the experimental questionnaire. The order of presentation of blocks of context plus target questions was randomised to minimise extraneous context effects.

Latency was measured automatically by the computer from the time the question was presented on the screen until a response was given. Respondents pressed a key for the next question to appear on screen. Respondents were not informed that response time was being measured.

After completing the computer questionnaire respondents were asked to fill in a short pen and paper questionnaire. This was the questionnaire dealing with knowledge and attitude structure factors. This was obtained from the researcher and completed in a room outside the computer booths.

6.3. RESULTS

This section presents the response results, that is the answers respondents gave to the questions. The focus is to look for differences across questionnaire context, indicating an effect of context on response. For each question three anovas were carried out.

The first anova looks for effects of the manipulated independent variables - questionnaire context and instruction. These are, essentially, factors which arise from the survey situation rather than factors which arise from the individual. The questions here are: Does the context in which the question is asked by itself have an effect on response? Do instructions on how to respond influence this effect?

The second anova looks for the effects of self reported knowledge on response. Three levels of knowledge (low, medium, and high) were derived from a 5-point knowledge scale. The question here is: Does the amount of prior knowledge a person has about an issue influence the effect of context?

The third anova deals with the 'attitude structure' factors -- importance and conflict. Importance was divided into important and unimportant from a 4-point scale. Conflict is whether the person's views are one-sided or conflicted. The question here is: Does the structure of the attitude influence context effects? For these factors, and for individual knowledge, we are not interested in main effects of these factors, since that only informs us about how individual factors influence positions on an issue. For these factors we are interested in interactions with context, as this informs us about the nature of context effects.

Separate anovas were conducted to determine the effects for each variable separate from others (this also allows more comparability with other studies where the various factors may not have been measured). However, attempts are made to look at whether the results are stable when the other factors are included. Because of the large number of anovas conducted, we might expect some results to be significant by chance. Significant results which are in line with predictions or are similar to other results, in the sense of forming a pattern of results, can more safely be regarded as non-chance results.

Those results which are anomolous, and unlike other results need to be treated more cautiously.

6.3.1 Manipulation checks

In general those who receive the 'think carefully', or 'slow', instruction should take longer to respond than those who receive the 'respond quickly', or 'fast', instruction. To check that the manipulation of instructions has worked mean latency was compared. Latency scores for all questions were added together to give a total latency score. This was compared across instruction. As expected, the slow instruction group took significantly longer to respond than the fast instruction group (t=4.33, p.000).

The pilot provides some evidence of the validity of the within subjects knowledge types and context type factors. The knowledge measure was repeated on the post experimental questionnaire and provides additional support for the validity of the knowledge type categorisation. Respondents have most knowledge about familiar issues, unfamiliar issues are least known and media issues come in between. Graphs giving mean knowledge, and summarising the importance, conflict and extremity of responses to issues are given in appendix 10.

6.3.2. RESULTS: RESPONSE EFFECTS.

First results which show an interaction between questionnaire context and thinking instruction will be presented. Second, results which show a main effect of questionnaire context will be presented, and finally results showing interactions between

questionnaire context and individual factors will be presented. Table 6.3, summarising the results for each question by the type of result is given below.

Table 6.3

Type of response effect by question

QUESTION	QUESTION TYPE		TYPE OF EFFECT		
Question name	knowledge type	context type	main effect context	context by instruct	context by knowledge/ attitude structure
ENVIRONMENT	Media	obvious	X		
CRIME	Media	obvious		X	
GENE	Unfamiliar	obvious	X		X
WHALE	Unfamiliar	obvious			
HOUSE	Familiar	obvious	X		
CLEAN	Familiar	obvious			
PORNOGRAPHY	Media	subtle			
EU	Media	subtle			
NEWTECH	Unfamiliar	subtle			X
OPERA	Unfamiliar	subtle			X
TEXT	Familiar	subtle			
FOOD	Familiar	subtle			

Looking first at interactions between questionnaire context and thinking instruction, only one question showed a significant interaction between these factors. This was the question about CRIME, a media issue with an obvious context. The question is given below (figure 6.1). The interaction between questionnaire context and instruction was significant at .042 (F=4.247). This level is reduced somewhat when

knowledge and attitude structure factors are included (.062
and .061).

Figure 6.1
CRIME Question

Question: Stricter punishment is necessary for many crimes. (5 point scale: agree/disagree)
Agree context: 'serious crimes'   Disagree context: 'trivial crimes'

In general there is slight agreement with this statement.
The interaction between questionnaire context and instruction
is that those who receive the slow instructions contrast with
the context whereas those who receive the fast instruction
assimilate. For those given the 'severe crimes' context, thus
promoting agreement to 'stricter punishment', those who
receive the slow instructions disagree more than those
receiving the fast instructions. For the 'trivial crimes'
context, promoting disagreement, those who receive the fast
instruction disagree slightly more than those receiving the
slow instruction.

Table 6.4
Mean Response to CRIME by
questionnaire context and instruction

	Agree		Disagree	
	Mean	SD	Mean	SD
FAST	3.84	1.14	3.52	1.44
SLOW	3.21	1.38	3.84	1.01

Figure 6.2
Mean Response to CRIME by
questionnaire context and instruction
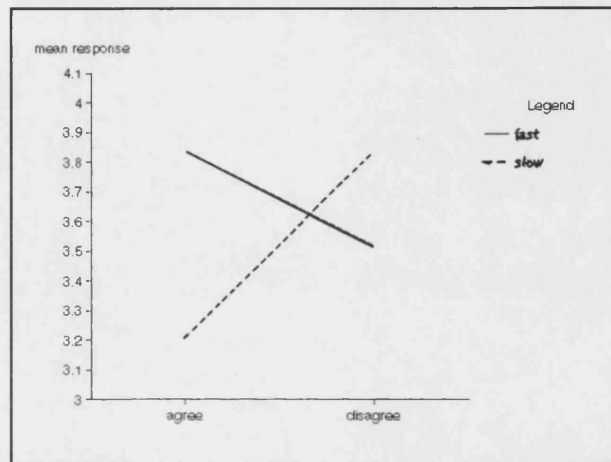
This is the only context effect for this question. There is
no main effect of context, nor are there any interactions with
attitude structure or knowledge factors.

Looking next at main effects of context, three questions show
some main effects. These are ENVIRONMENT, HOUSE, and GENE.
All three are issues with an obvious context. Each comes from
a different knowledge category.

The strongest, and most stable main effect of questionnaire
context is for ENVIRONMENT, a media issue with obvious
context.

Figure 6.3
ENVIRONMENT question

Question: New jobs should be created even if this sometimes causes damage to the countryside. (5 point scale: agree/disagree)
Agree context: 'Job creation' Disagree context: 'Environment protection'

The main effect of questionnaire context is significant at
.007 (F=7.693). This is relatively stable with slightly
reduced significance when knowledge, and importance and
conflict are added (.01 and .016 respectively).

Table 6.5
Mean response to ENVIRONMENT by
questionnaire context

	Agree	Disagree
Mean	2.28	2.84
S.D.	1.01	1.02

In general there is a tendency to disagree with this
statement. The main effect of questionnaire context is a
contrast effect. So those who receive the 'job creation'
context, which should lead to more agreement, disagree more
with the 'jobs at the expense of countryside' proposition than
do those who receive the 'environment protection' context,
which should lead to more disagreement.

This is the only significant context effect for ENVIRONMENT.
There is a main effect of conflict, but this simply shows that
those who are one-sided disagree more with this statement than
those who are conflicted.

For HOUSE, a familiar issue with obvious context, there is a
main effect of questionnaire context (F=3.869, sig .052).
However, this is somewhat unstable acros the different ANOVAs.
The effect is strengthened when knowledge joins the ANOVA

(.041) and disappears when conflict and importance are included (.103).

Figure 6.4
HOUSE question

Question: Students in higher education are adequately
            housed.
            (5 point scale: agree/disagree)

Agree context: 'homeless in London'
Disagree context: 'mortgage relief'

Table 6.6
Mean response for HOUSE by
questionnaire context

	Agree	Disagree
Mean	2.88	2.39
S.D.	1.15	1.34

Overall there is slight disagreement with this statement. The effect of questionnaire context is an *assimilation* effect. Those given the 'house buying' context, prompting disagreement, disagree more than those given the 'homeless' context, prompting agreement.

This is the only context effect for this question. All the effects of knowledge and attitude structure effects are main effects.

For GENE, an unfamiliar issue with obvious context, there is a slight, but stable, main effect of context. This is significant at .071 (F=3.326), a value which remains stable

across the different analyses. Mean response differs slightly across context condition.

Figure 6.5
GENE question

Question: Genetic engineering research is adequately controlled. (5 point scale: agree/disagree)
Agree context: 'good technology' Disagree context: 'bad technology'

There is a general tendency to slightly disagree with this statement. This slight effect of questionnaire context is a *contrast* effect. Those who receive the 'benefits' context, promoting agreement with the adequacy of control of genetic research, disagree more than those who receive the 'risks' context, which promotes disagreement.

Table 6.7
Mean response to GENE by
questionnaire context

	Agree	Disagree
Mean	2.18	2.56
S.D.	0.99	1.07

However, for this question, questionnaire context also interacts with knowledge and attitude structure factors. The interaction of questionnaire context with knowledge (F=7.742, sig .003) is that for those with little or no

knowledge there is little difference between questionnaire
context. However, for those with a moderate amount of
knowledge, there is a contrast effect such that those who
receive the 'agree' context disagree more than those who
receive the 'disagree' context. Those with most knowledge
also follow this pattern, with the 'disagree' context higher
than the 'agree' context. However, due to the small number of
those with much knowledge this needs to be interpreted
cautiously.

Table 6.8
Mean response to GENE by
questionnaire context and knowledge

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
High knowledge	2.17	1.47	6	4.50	0.71	2
Mid knowledge	1.68	0.84	22	2.56	1.15	16
Low knowledge	2.71	0.72	21	2.44	0.95	32

The result is similar when those who have much knowledge are
combined with those who have some knowledge. This simplifies
the interpretation somewhat. Those with little knowledge show
little difference between questionnaire context. However,
those with at least some knowledge show more of a contrast
effect between questionnaires. Those in the 'disagree'
context disagree less than those in the 'agree' context.

Table 6.9
Mean response to GENE by
questionnaire context and knowledge
(knowledge split into two groups)

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
Some knowledge	**1.79**	0.99	28	**2.78**	1.26	18
Low knowledge	**2.71**	0.72	21	**2.44**	0.95	32

Figure 6.6
Mean response to GENE by
questionnaire context and knowledge
(knowledge split into two groups)
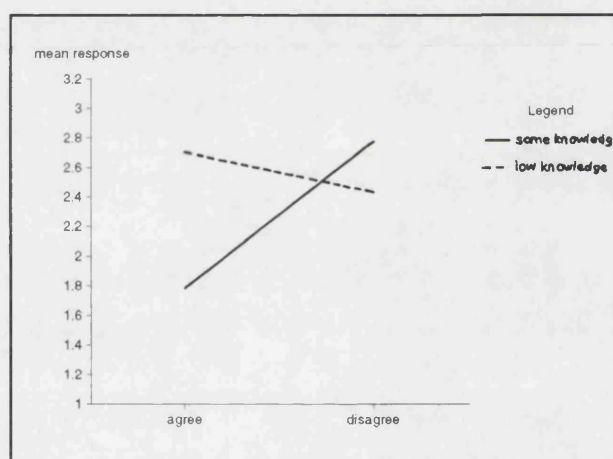
For the attitude structure factors there are a number of significant interactions. Questionnaire context interacts with importance (F=6.389, sig .013) and interacts slightly with Conflict (F=3.46, sig .066). The thinking instructions also interact with with Conflict (F=6.11, sig .015) and with Importance (F=4.60, sig .035), though these are of less interest in looking at context and response (Tables for these are given in appendix 11).

The interaction between importance and questionnaire context is such that for those who think the issue is unimportant, there is little difference in response between the two questionnaires. However, those who think the issue is important contrast more; those who receive the 'disagree', context disagree less than those who receive the 'agree' context.

Table 6.10
Mean response to GENE by
questionnaire context and importance

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
Important	**1.79**	0.92	28	**2.56**	1.23	25
Unimportant	**2.71**	0.85	21	**2.56**	0.92	25

Figure 6.7
Mean response to GENE by
questionnaire context and importance
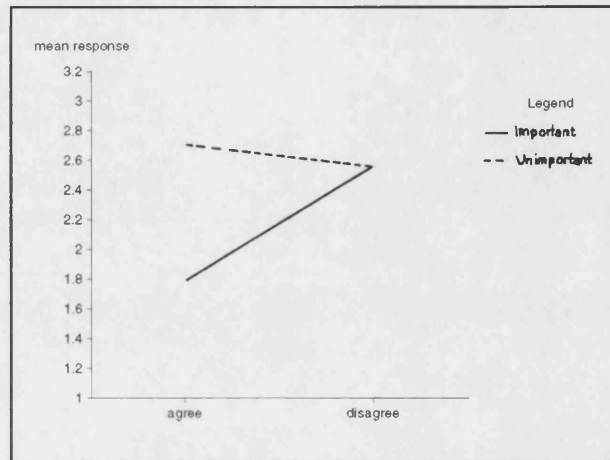
The slight questionnaire context by conflict interaction is
such that the one-sided people differ little between
questionnaire context. There is a slight tendency for those
who are conflicted to contrast more; those given the agreement
context disagree more than those given the disagree context.

Table 6.11
Mean response to GENE by
questionnaire context and conflicted

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
one-sided	2.06	1.18	16	2.19	1.28	16
conflicted	2.24	0.90	33	2.74	0.93	34

Figure 6.8
Mean response to GENE by
questionnaire context and conflicted
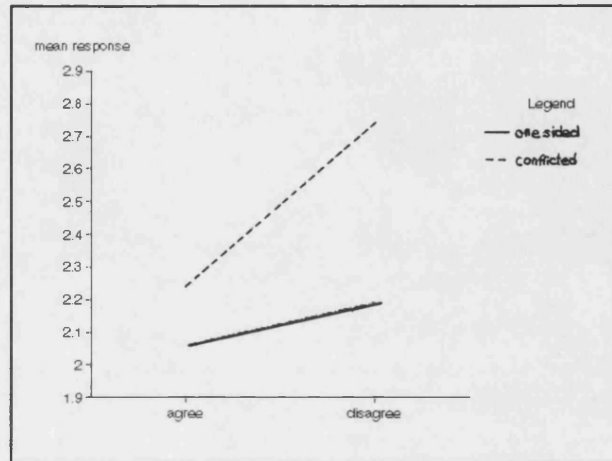
Other questions which show interactions between questionnaire context and either knowledge or attitude structure factors are the questions on new technologies, NEWTECH, and the OPERA. These two issues are both unfamiliar issues with subtle contexts.

For NEWTECH, an unfamiliar issue with subtle context, while there is no main effect of context, there are interactions between context and knowledge and between context and conflict.

Figure 6.9
NEWTECH question

Question: Moves to introduce new technologies, such as
           'virtual shopping', into daily life should be
           resisted.
           (5 point scale: agree/disagree)

Agree context: 'historical viewpoint and isolation'
Disagree context: 'looking forward not back'

The interaction between knowledge and questionnaire context is significant at .026 (F=3.793). This is stronger when the attitude structure factors added (.014). For those with little knowledge there is little difference across context (i.e. no effect of context). Those with some knowledge tend to assimilate to the context and those with high knowledge tend to contrast with the context.

Table 6.12
Mean response to NEWTECH by
questionnaire context and knowledge

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
High knowledge	**2.10**	1.60	10	**2.91**	1.70	11
Mid knowledge	**3.33**	1.37	18	**2.43**	0.94	14
Low knowledge	**2.29**	0.85	21	**2.52**	1.05	25

Figure 6.10
Mean response to NEWTECH by
questionnaire context and knowledge
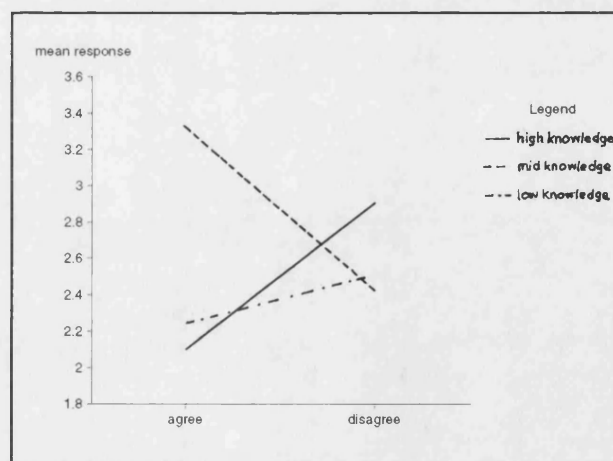
There is also an interaction between questionnaire context and Conflict (F=5.785, sig .018. This is stronger when knowledge is added .005). Those who are one-sided assimilate more to the

context than those who are conflicted. Thus for the disagree
context, one-sided people disagree more than those who are
conflicted. For the agree context, one sided people agree
more than those who are conflicted. Alternatively, one might
say that conflicted people contrast more.

Table 6.13
Mean response to NEWTECH by
questionnaire context and conflicted

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
one-sided	**3.05**	1.61	20	**2.11**	1.28	18
conflicted	**2.34**	1.01	29	**2.84**	1.05	32

Figure 6.11
Mean response to NEWTECH by
questionnaire context and conflicted
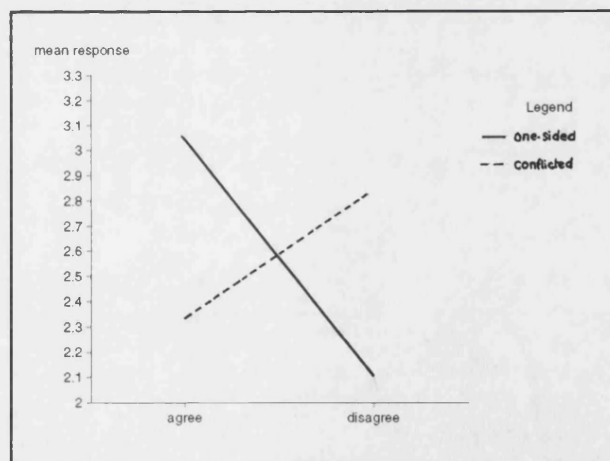
For OPERA, also an unfamiliar issue with subtle context, there
is no main effect of context but there is an interaction
between questionnaire context and importance.

Figure 6.12
OPERA question

Question: The government should provide more support
           for the opera.
           (5 point scale: agree/disagree)

Agree context: 'national heritage'
Disagree context: 'élitism'

Table 6.14
Mean response to OPERA by
questionnaire context and importance

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
important	**4.00**	1.08	13	**2.80**	1.79	5
unimportant	**2.39**	1.02	36	**2.93**	1.10	45

Figure 6.13
Mean response to OPERA by
questionnaire context and importance
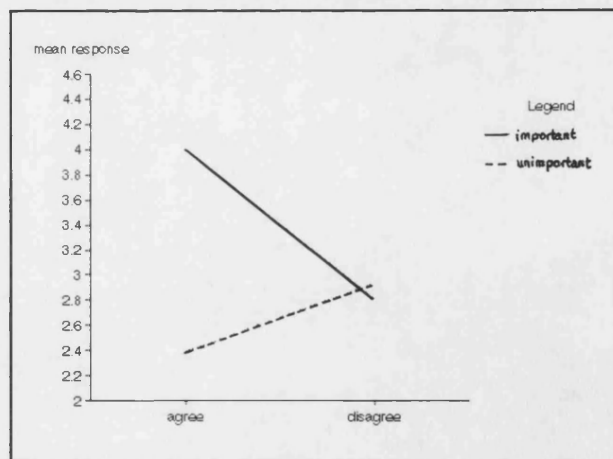
This interaction effect (F= 6.569, sig .012) is that for those

given the 'élitist' context, promoting disagreement, there is

little difference between those who think it unimportant and

those who think it important. For those given the 'heritage' context, promoting agreement there is a large difference between those who think it unimportant and those who think it important. Those who think it important assimilate to the context more while those who think it unimportant do not. However, one must be cautious with this interpretation, because of the small number of people who think this issue important. In fact it is the least important issue as well as the one for which people know least. It is possible that those who do care about it understand the 'heritage' context more readily than those who do not care and are thus more influenced by it.

The other questions used here show little or no context effects of any type. These include CLEAN, a familiar issue with obvious context; TEXT and FOOD, both familiar issues with a subtle contexts; WHALE, an unfamiliar issue with obvious context; and PORNOGRAPHY and EU, both media issues with subtle contexts.

6.3.3 SUMMARY OF RESULTS

The strongest main effect of context is for ENVIRONMENT, a media issue with obvious context. HOUSE, a familiar issue with obvious context, also has a significant main effect of context, though not when attitude structure factors are taken into account. For GENE, an unfamiliar issue with obvious context, the main effect of context tends towards significance. For ENVIRONMENT and GENE the effects are

contrast effects. For HOUSE the effect is an assimilation effect.

There is only one question for which questionnaire context interacts with instruction, this is for CRIME, a media issue with obvious context. Those given the fast instruction assimilate more, those given the slow instruction contrast more. Table 6.15 below gives the number of effects broken down by knowledge and context type.

There are also a number of context effects which stem from interactions with knowledge and attitude structure factors. Three of the four unfamiliar questions have this type of effect: GENE (obvious context), with interactions with knowledge, conflict, and importance. OPERA and NEWTECH (both subtle context), the former with an interaction with importance, the latter with an interaction with knowledge and with conflict.

Table 6.15
Number of context effects by type of effect
And knowledge + context type

	MEDIA OBVIOUS	MEDIA SUBTLE	UNFAM OBVIOUS	UNFAM SUBTLE	FAM OBVIOUS	FAM SUBTLE	TOTAL
QAIRE or QAIRE + INSTRUCT	2/4	0/4	1/4	0/4	1/4	0/4	4/24
QAIRE + ATTITUDE or KNOWL	0/6	0/6	3/6	3/6	0/6	0/6	6/36
TOTAL	2/10	0/10	4/10	3/10	1/10	0/10	10/60

In terms of general response effects due to questionnaire factors - there is little difference across knowledge type although perhaps the strongest effects are to be found among the media issues. Table 6.16 below illustrates this. Response effects due to a combination of questionnaire and personal factors are more common for the unfamiliar issues.

Table 6.16
Number Of context effects by
Type of effect and knowledge type

	MEDIA	UNFAMILIAR	FAMILIAR	TOTAL
QAIRE or qaire + INSTRUCT	2/8	1/8	1/8	4/24
QAIRE + ATTITUDE OR KNOWL	0/12	6/12	1/12	6/36
TOTAL	2/20	7/20	1/20	10/60

In terms of the type of context it is clear that all the response effects due simply to questionnaire & instruction factors come from those questions which have an obvious context. However, where the effect comes from a combination of questionnaire and personal factors the type of context is less important. Table 6.17 below illustrates this.

Table 6.17
Number of context effects by
Type of effect and context type

	OBVIOUS	SUBTLE	TOTAL
QAIRE or qaire + INSTRUCT	4/12	0/12	4/24
QAIRE + ATTITUDE OR KNOWL	3/18	3/18	6/36
TOTAL	7/30	3/30	10/60

6.4 DISCUSSION

First I will discuss the results in terms of the independent variables of context and instruction, and then in terms of interactions of context with the individual factors. I then want to look at the factors of knowledge type and context type to see if these distinctions add anything to the interpretation of the context effects.

The first thing to note about these results is the variation in context effects over questions. Some questions show effects, some do not. Some show main effects, some interactions. Some show assimilation effects, some contrast, some both. The question is whether we can determine factors which are influential in producing these different effects, or is it simply a matter of different questions producing different effects?

Only one question showed an effect of thinking instruction. This suggests that the manipulation of thinking is likely to be only a minor factor in producing context effects. The association of contrast with more complex thinking is an established idea in the literature (Martin & Harlow, 1992; Strack, 1992). The one question which showed an effect of thinking instruction in the present study supports the idea that taking more time, or thinking carefully, is more likely to produce contrast, whilst assimilation is more associated with quick, less thoughtful, response. Although only one question showed this relation, this is an important finding; it shows, within a survey context, that a simple manipulation of thinking can alter response.

In terms of the individual factors -- knowledge, importance, and conflict -- there is no overall pattern to the results. One question which shows an overall effect can be qualified by these factors. This is GENE where the contrast effect is stronger for those with more knowledge, for those who think the issue important, and for those who are conflicted. Other issues show some effect of these factors, but have no overall context effect (OPERA, NEWTECH), other questions show context effects which are not qualified by these factors (ENVIRONMENT, CRIME, HOUSE), although with HOUSE the overall effect is eliminated when these factors are included in the ANOVA.

The effects of conflict and importance for GENE are in the expected direction. However, the effects for importance and conflict are independent effects rather than an interaction

between the two. Tourangeau et al (1989a, 1989b) found that response effects were stronger, though not significantly so, for those who were both conflicted and thought the issue important; conflict produced some effect on its own, but importance did not.

There may be several reasons why no interaction effects between the two attitude structure factors and questionnaire context appeared in this study. First, perhaps the particular issues used do not lend themselves to such effects. Second, perhaps the sample size was too small to identify such interactions.

Where knowledge interacts with context, for GENE and NEWTECH, it is those who have more knowledge who show a larger effect. This is somewhat at odds with studies looking at other aspects of attitudes, where for example those with more knowledge are more resistant to change (Wood, 1982). And, one might expect those with more knowledge have access to other information which would attenuate the effect of context (Feldman & Lynch, 1988). However, the above results are consistent with Tourangeau et al's (1989a, 1989b) reasoning behind why those who see the issue as important show more effect. They suggest that those who see the issue as important have elaborated attitude structures which include links to related issues; this also implies a greater degree of knowledge about an issue. Because they are able to see links between issues, they are more likely to be affected by context. That knowledge has an effect in this study suggests this is the

case, at least for some issues, but the fact that importance has an independent effect suggests that it should not simply be equated with knowledge on an issue.

The more puzzling effects of importance come with OPERA, although some cell sizes are too small to place much confidence in this result.

Thus, so far, a number of different effects have been seen to occur for these questions with no clear reasons why some questions show one type of effect whilst others show another type. It may well be the case that the nature of individual questions is important in determining whether and what type of context effects occur. This would be an unfortunate result for those trying to design questionnaires, since it would mean that individual questions would always need separate investigation. However, question and context type have also been examined in this study, and it may well be that these factors can add to our understanding of context effects.

First, it is important to note in discussing the factors of knowledge and context type used in this study that the results can only be suggestive rather than conclusive in regards to the effect of these factors. Because different questions are used to assess these factors, we cannot rule out the possibility that particular questions are responsible for effects rather than any higher order grouping factor. A stronger test of the knowledge factor would be possible by using more questions from each category, and a stronger test

of context type would be possible by comparing the same target with different types of context. However, this was beyond the scope of the present experiment. Having said this, the present results do suggest that these factors may be important in the impact of context on response.

All the effects which come from the situational factors, questionnaire and instructions, are for questions with obvious contexts. This suggests that the relatedness of context to target is important in determining context effects. It seems for overall context effects, the context needs to be conceptually related to the target. This fits well with suggestions in the literature that the diagnosticity of previous questions is important, that is, the respondent must be able to perceive some relation of context and target (Feldman & Lynch 1988). This relationship cannot simply be determined by the correlation between context and target questions. All 4 issues which show task related context effects are among the highest rated 6 issues in terms of relatedness of context to target; all had mean ratings near or above 'somewhat related'. The correlations between target and context are generally no higher for the issues where context effects appear than for issues where they do not (see appendix 12).

Knowledge type seems less important for overall effects, although it might play some role. The strongest overall effect comes on a media issue, with the other media issue showing an interaction with instruction. The main effect on

the unfamiliar issue is weak and the one main effect found at a familiar issue is unstable. Thus whilst the evidence for context differences due to situational factors is weak for knowledge type, there is some suggestion that media issues may be most susceptible to overall context effects, with unfamiliar and familiar issues less susceptible.

For familiar issues there may be less effect for an obvious context because people have more stable ideas about the issue through experience, thus, although they may recognise the implications of context for target it is less influential because they have more other easily accessible information to draw on. That unfamiliar issues are less affected is somewhat more difficult. Others have shown context effects for obscure or fictitious (and thus presumably unfamiliar) issues (Strack, 1992; Tourangeau & Rasinski, 1988). However, these kinds of context effects were most likely interpretational; the previous context provided a kind of definition of the fictitious issue allowing comprehension of the fictitious issue in terms of the previous question. In this study the context items are unlikely to have aided comprehension, at least not in such a straightforward way. Rather, it may be that for these unfamiliar issues the context is less diagnostic for the target simply because the issue is unfamiliar. Whereas for media issues respondents have some knowledge about the issue, and also about how that issue relates to other issues, for unfamiliar issues they may have less of both these types of knowledge, and so be less

influenced by the context because they cannot so readily relate the target issue to the context issue.

In terms of the individual factors, for the media and familiar issues most of the effects of knowledge and attitude structure factors are main effects. That is, their influence on the answer is not dependent on the context. Rather, knowledge and attitude structure factors can be seen as related to positions on the issue, regardless of context.

For the unfamiliar issues, however, attitude structure and knowledge factors are more important. Three of the four issues have interactions of context with these factors. But, it is also the case that the type of context, obvious or subtle, distinguishes less between these interaction effects for unfamiliar issues. Both of the unfamiliar issues with a subtle context show interactive context effects.

In the case of GENE, where we do get some overall context effect, the effect is stronger for those who are more knowledgeable, for those who think the issue more important, and for those who are conflicted. The above suggestion that the context at unfamiliar issues is less diagnostic because of a lack of links to other issues would be supported by this. Those who are more knowledgeable, or who see the issue as more important, are more likely to perceive links to the previous context and thus to be influenced by it. A similar situation exists for NEWTECH, where those with little knowledge are unaffected by context, but those who have more knowledge are

affected, though in different directions depending on the level of knowledge.

But why do we find interaction effects for unfamiliar issues and not for media issues? Tourangeau et al's studies (1989a, 1989b) showing greater context effects for those who were conflicted and saw the issue as important used mainly media issues. It is possible that, at least in part, the nature of the sample might be responsible for these differences. Tourangeau et al used a general population sample; this study used a relatively homogenous student sample. The degree of exposure to topical media issues might be expected to be more uniform in the student sample than in a general population sample. This similarity of exposure to issues may serve to make the individual's structure of the issue less influential. For unfamiliar issues on the other hand, there may be more of a variety of exposure to issues, some people may be opera buffs for example, or read a lot about genetic engineering, whereas others may not. The fact that these issues are not such topical media issues means that exposure to them is more likely to be driven by personal interest rather than simply from reading the newspaper or watching the television news. In this way individual structuring of these issues may be more influential because the range of exposure is greater.

The implication here also is that for media issues people may, generally, have more alternative diagnostic inputs available. The subtle context is not very diagnostic; when people have alternative, more diagnostic, inputs more readily accessible,

they may be more likely to use these inputs than to use an available but weakly diagnostic context (Feldman & Lynch, 1988). However, for unfamiliar issues, even though the context is only mildly diagnostic there may not be alternative, more diagnostic information easily accessible, and thus, the less diagnostic, but easily available information is used, especially among people who are capable of recognising its diagnosticity (those with more knowledge).

Apart from what factors influence whether there is an effect of context, there is also the question of the direction of the effect -- assimilation or contrast. What determines the direction of an effect? Most of the effects here are contrast effects. However, there are also assimilation effects. And, sometimes both occur for the same question depending on the interaction of various factors. Context type does not seem to discriminate -- there are both assimilation and contrast effects for both types of context. Knowledge type may discriminate to some extent. For familiar issues only assimilation effects occur. For media and unfamiliar issues both types of effects occur, though for the effects which are solely based on context, only contrast effects are found. The direction of the effect is thus difficult to explain by the various associated factors, and it may be that it is dependent on the particular question.

Tourangeau et al, looking at a number of issues generally found assimilation effects. In this study we find more contrast effects. One reason why more contrast may occur in

this study is that with more knowledgeable respondents they may either be more motivated to respond well, or be more capable of linking issues. Contrast is suggested to occur when people are aware of the previous context, and are motivated to overcome this biasing influence (Martin, 1986). Knowledgeable respondents may be more likely to be both interested enough to want to provide a good answer, and knowledgeable enough to recognise an obvious context (Bickert, 1993). Looking at response order effects, Krosnick & Alwin (1987) found greater effects for those with lower levels of education and suggested this was due to their greater use of satisficing strategies to provide easy answers. Our respondents are all relatively well educated, and thus may satisfice less. Thus, for ENVIRONMENT, they may quickly recognise the biasing influence of previous questions, since they are so obvious, and adjust for this influence. At GENE, those who are more knowledgeable show greater contrast effect; likewise at NEWTECH, with a subtle context, it is those who are most knowledgeable who contrast whilst those with some knowledge assimilate. These results suggest that greater knowledge may lead to contrast. With CRIME, however, we have both assimilation and contrast effects, although the context effect in the fast, 'assimilation', group is relatively small. Perhaps for this particular issue and context it takes more time to recognise the biasing nature of the context than it does for the ENVIRONMENT issue, hence, the contrast effect occurs only for those who are instructed to think carefully. Or, perhaps it is easier to 'call up' opposing implications for ENVIRONMENT; for ENVIRONMENT, both sides of the issue are,

to some extent, included in the question, as it mentions both 'job creation' and 'countryside damage', whereas CRIME does not. Certainly knowledge, at some level, seems to be implicated in these effects. However, at the familiar issues, where one would expect most knowledge, we get an assimilation effect, though admittedly an unstable one. Clearly further research on these factors would be relevant.

In summary these results have the following implications for context effects at attitude questions:

1. Manipulating the amount of thinking at a question can influence the direction of context effects in the way predicted by an assumption that contrasting takes longer than assimilating. However, the effects are likely to be limited. Examinations of latencies in the next chapter offer some further suggestions about the effects of thinking time on context effects. Investigating what kind of questions, and perhaps respondents, are most likely to be influenced by thinking manipulations would be useful.

2. The individual factors of importance and conflict are implicated in some context effects, such that those who see the issue as more important show more effects and those who are conflicted show more effects. However, the results do not support Tourangeau et al's findings of an interaction between these two factors. It is not clear whether this is due to limitations in the present study or whether the type of

questions involved is important. Further research investigating these two factors would be interesting.

3. Individual knowledge is implicated in some context effects. Those with more knowledge show more effect. Individual knowledge is likely to be important in context effects where the variation in knowledge is greater.

4. Knowledge type and context type are likely to be important in context effects. As expected, familiar issues show least effects. Media issues show effects only in an obvious context. Unfamiliar issues show effects which interact with the individual factors of knowledge and attitude structure. The distinction between issues based on an aggregate knowledge level may be important in understanding the nature of context effects. This issue, and others raised by this study, are discussed further in the concluding chapter.

CHAPTER 7 - EXPERIMENT 3 -- PART 2

ATTITUDES AND RESPONSE LATENCY

7.1 INTRODUCTION

As argued in the previous chapter, think aloud does not recommend itself as a method for tracing processing when thinking time is an important issue. Therefore another method to gain information about the cognitive processes underlying context effects was sought. Another method commonly used in cognitive psychology is the measurement of response latencies. This has the advantage of being a non-disruptive measure, although it gives less direct information on processing.

Fazio (1990; also Dovidio & Fazio 1992) documents the more recent use of latency measures in social psychology. He reviews several areas in which latency measures have been used to examine construct formation, processing efficiency, and associative strength in memory. This latter use of latency measures is one which may be most useful for survey methodology.

Context effects are linked to the accessibility of information, with the context making particular information accessible by virtue of its recent use. Other factors, such as the chronic accessibility of an attitude or the easy accessibility of other information also play a part. But again, the key is accessibility. Latency is often used as a

measure of accessibility. The reasoning being that the more accessible a piece of information is the quicker it will be reported. Availability is assumed to result in faster response time. Thus, factors which increase the availability of the response should result in quicker latencies.

Fazio and colleagues (Fazio, Herr, & Olney, 1984; Fazio, Powell & Herr, 1983; Fazio, Sonbonmatsu, Powell, & Kardes, 1986; Fazio & Williams, 1986) have done a considerable amount of research which uses response latency as a measure of attitude accessibility. They show that this measure of accessibility relates to attitude-behaviour correspondence. Similarly, Basili and Fletcher (1991) use latency measures to examine attitudes and non-attitudes. They asked respondents two questions about the issue of employment quotas and found that those who responded differently on the second question, that is changed their opinion, who they term 'movers', took longer than non-movers to respond to the quotas question. They suggest this shows that non-movers have more accessible, and thus perhaps more crystallised attitudes.

Fazio (1990) also looks at the association between category labels and members, using latency as a measure of that association, the faster the latency, the stronger the association. He suggests that because this measure correlates well with a number of other measures of association, latency provides a good measure of associative strength in memory. Judd, Drake, Downing & Krosnick (1991) provide similar evidence using related attitude issues, where they find that

responses to an issue are quicker when asked after a related rather than an unrelated issue. Tourangeau, Rasinski & D'Andrade (1991) also show that response latency may be used as a measure of associative strength; they show that when targets follow items drawn from related clusters response latencies are quicker.

Quicker response time has also been associated with a number of attitude factors, some of which have also been related to response effects. Important attitudes have been associated with quicker response (Bassili, 1993; Krosnick, 1989; Roese & Olson, 1994) as have more extreme attitudes (Fazio & Williams, 1986; Judd & Kulik, 1980). Presumably because attitudes with these characteristics are more accessible. In the present study one might also expect one-sided attitudes to be related to quicker response as opposed to conflicted attitudes. High knowledge may also be associated with a quicker response time, however this may be influenced by whether this is associated with one-sided or conflicted attitudes.

Attitudes which are automatically activated, with quick latencies, tend to be those of which one has direct experience (Fazio et al 1986). Thus, one might expect that more familiar issues will be associated with quicker response times than unfamiliar issues.

A simple association of response time with attitude accessibility may not however be warranted in a survey setting. As Martin (1986) has pointed out, processing

objectives are also important in determining response. Contrast effects are associated with greater cognitive effort on the part of the respondent. Greater cognitive effort may result in a longer latency. In this case, we might expect that contrasting is associated with slower latencies. Furthermore, a related previous context may actually result in slower latencies if respondents have to decide whether this information should be used in forming a judgement.

7.2 RESULTS

Latencies were converted to logs for analysis. Raw latency scores tend to be skewed, with a long tail of slower latencies (perhaps due for example to respondents being temporarily distracted, or searching for the correct key to press). Thus, for latency scores the mean tends to be a poor measure of central tendency. A logarithmic transformation is commonly used to bring the tail closer to the centre of the distribution so that the mean is a better measure of central tendency (Fazio, 1990).

The results for latency will be presented in terms of the question they address. Firstly differences across instruction will be compared. Secondly, how latency differs across different knowledge and context types will be examined. Thirdly, the effect of previous context on latency will be examined. Fourthly, individual knowledge and attitude structure factors will be examined. Finally, latencies for assimilation and contrast will be examined.

7.2.1 Latency across instructions.

To give some idea of the real times involved in responding to questions, table 7.1 gives the mean raw latency scores for each question, across fast and slow instructions. Table 7.2 gives mean log latency for each question across instruction. The results of 2-way Anovas, Questionnaire context by instruction, are also given in table 7.2, showing the significance of the difference in latencies between fast and slow instructions for each question.

Table 7.1
Mean Raw Latency Scores by Instruction

question	fast		slow		min	max
	Mean	S.D.	Mean	S.D.		
Envirnoment	7.69	4.0	11.27	7.1	2.75	36.75
Crime	5.78	3.0	7.61	4.0	2.31	20.05
Pornography	6.54	3.2	9.42	6.1	2.69	32.30
EU	7.58	3.1	10.25	7.2	3.03	44.00
Gene	7.17	3.8	8.62	6.4	2.58	41.85
Whale	12.48	5.7	14.67	7.1	4.78	35.31
Opera	5.29	1.6	6.92	3.1	1.97	15.82
Newtech	9.22	3.6	10.84	5.9	2.04	33.01
House	6.51	2.8	8.18	5.5	2.69	29.44
Clean	5.19	1.7	7.14	3.2	2.36	15.00
Food	8.21	2.9	10.28	5.4	3.35	28.73
Text	6.95	3.0	8.29	4.9	3.52	33.89

Table 7.2
Mean log latency by instruction

Question	fast	slow	sig level
Environment	1.94 (.44)	2.27 (.54)	.002
Crime	1.64 (.47)	1.91 (.50)	.007
Pornography	1.78 (.42)	2.08 (.56)	.003
EC	1.95 (.38)	2.16 (.55)	.032
Whale	2.43 (.47)	2.57 (.50)	.155
Gene	1.85 (.48)	1.99 (.53)	.176
Opera	1.62 (.31)	1.84 (.43)	.003
newtech	2.15 (.37)	2.26 (.49)	.209
house	1.79 (.40)	1.94 (.53)	.104
clean	1.60 (.30)	1.91 (.44)	.000
text	1.86 (.39)	2.00 (.45)	.107
Food	2.05 (.34)	2.23 (.43)	.019

We would expect that those given the instruction to answer quickly would have shorter latencies than those given the instruction to take their time. And, indeed when we compare overall latency this is the result we get (reported in Chapter 6). For the individual questions all the results are in the same direction, the fast instructions produce quicker latencies than the slow instructions. However, the extent of the difference in the effect of instruction is not uniform across the target questions.

7.2.2 Latency and knowledge and context types

To look at how latency varies across knowledge and context types a combined latency score was computed. As reading time was included in the latency, this had to be deducted for direct comparisons to be made, as questions were of different length. Reading time was calculated for each question (a set amount of time per word) and subtracted from the latency. These adjusted means, which allow comparisons between

questions, are given in appendix 13. Although this measure may have some problems (eg. a set amount per word overlooks variation in reading time) it also has some benefits over other methods. For example, starting latency measures from the end of a question (ignoring reading time) neglects any thinking done while reading (eg some people may take more time than others to think about the meaning of a question while they are reading). Latencies were then summed for each knowledge by context type category and the log taken. Table 7.3 gives mean log latencies by knowledge and context type.

Table 7.3
Mean log latency by knowledge and context types

	obvious	subtle	overall
unfamiliar	1.91	1.66	1.79
media	1.64	1.64	1.64
familiar	1.56	1.58	1.57
overall	1.70	1.63	

A manova was performed with knowledge type and context type as within subject factors and instruction as the between subject factor. This showed significant main effects for both knowledge and context types (F=19.55, sig .000 and F=6.59, sig .012 respectively) and a significant interaction between them (F=6.52, sig .002).

As can be seen the unfamiliar issues take longest to answer, and the familiar issues take least time, with media issues in between. This is particularly so in the obvious context; differences in the subtle context are slight. For media and

familiar issues the latencies are similar accross obvious and subtle contexts but for the unfamiliar issues there is a difference between context type with longer latencies for the obvious context than for the subtle context.

However, it could simply be that the question showing the largest response time (WHALE) is simply a bad question (eg difficult to understand) and therefore increases response time.

It is interesting to note that the fastest response time occurs for an unfamiliar question (OPERA). This question is also seen as the least important issue. So perhaps we have an exception to the idea that unfamiliar issues take longer. Perhaps if the issue is seen as particularly unimportant, then it is deemed not worth thinking about and so people respond more quickly. Further work investigating this relationship more thoroughly may be profitable.

There is also a slight interaction between knowledge type and instruction (F=2.72, p =.068) (see table 7.4 below). This is such that the largest difference in instruction comes from the media issues, followed by familiar issues, and the smallest difference occurs for unfamiliar issues, confirming the question by question results of the last section. What these results perhaps suggest is that for both familiar and media issues one can respond quickly, for unfamiliar issues, however, it is difficult for people not to think -- to respond off the top of their head. Note that the mean for unfamiliar

issues in the fast instructions is virtually the same as for familiar issues in the slow instructions. It may also show, somewhat, that for the familiar issues, it is more difficult for people to take time. They don't have as much to think about since they are more familiar with the issue and perhaps have clearer views on it. Thus in the fast condition media and familiar issues have more similar latencies, whilst in the slow condition media and unfamiliar issues are more similar. It seems media issues (often the type of issue questioned in surveys) are more variable in the amount of time that can be spent thinking about them.

Table 7.4
Mean log latency for knowledge type by instruction

	fast	slow
familiar	1.44	1.70
media	1.47	1.82
unfamiliar	1.69	1.89

7.2.3 Prior context and latency.

It has been suggested, and in some cases shown, that a related prior context makes latency to a subsequent related issue quicker (Judd et al, 1991). This question cannot be addressed directly in this study because a strong test of this would require the use of different contexts for the same question. To some extent, this issue is addressed above (Table 7.3) where it is shown that latencies to targets in an obvious context are slower or no faster than targets in a subtle context. This question might be addressed further by comparing context and target latencies. One might expect that

target questions following an obvious context will be faster than the context latency, since this context may facilitate response to the target. However, this supposition needs to be treated cautiously, since context may differ from targets in other ways which might make them answered more slowly (or more quickly). One obvious problem is that the context at unfamiliar target questions may be more familiar than the target, and answered more quickly for this reason; the reverse may be the case for familiar targets. However, across all context questions, there is no reason to assume that the context questions differ systematically in any way other than that some are obvious and some are subtle, and thus, over all questions one might expect that target latencies should be faster than context latencies in obvious contexts but not in subtle contexts.

To address this question the latency scores for the context questions were combined and the mean taken and compared to the means for target latency (note that context latencies were also adjusted for reading time). A manova was then performed with knowledge type, context type, and context-target as within subject factors, and instruction as a between subject factor. Over all questions the latency for target questions is faster than the latency for context questions (F=5.77, sig .018). There is an interaction between context type and context-target such that in obvious contexts there is little difference between context and target latencies but for subtle contexts target latencies are faster overall than context latencies (F=19.25, sig .000).

Table 7.5
Mean log latency for context and target questions by
obvious /subtle context type

	context latency	target latency
obvious context	1.68	1.70
subtle context	1.75	1.63

There is also an interaction with knowledge type such that overall for media and unfamiliar issues there is little difference between context and target latencies, but for familiar issues target latencies are faster than context latencies (F=11.41, sig .000). This latter result suggests that for familiar issues the targets are more familiar than the context and thus the targets are answered more quickly.

Table 7.6
Mean log latencies of context and target
by knowledge type

	context latency	target latency
familiar issues	1.77	1.57
media issues	1.64	1.64
unfamiliar issues	1.74	1.79

The idea of target latencies being no different than context latencies in an obvious context is contrary to expectations, but fits with the earlier finding of overall target latencies being slower or no different in obvious contexts. It is perhaps not the case that an unrelated context speeds up response, but rather that a related context can slow it down.

People may think about how the previous context relates to the present question and take more time over it.

7.2.4 Individual knowledge and latency

In general individual knowledge makes little difference to latency. Those who know more about an issue do not respond more quickly than those who know little. The exception to this is the opera question where those who know more about the issue respond more quickly (F=9.75, sig .000).

Table 7.7
Mean log latency for OPERA by knowledge

	Mean	SD	N
High knowledge	**1.39**	0.39	10
Mid knowledge	**1.70**	0.31	21
Low knowledge	**1.79**	0.39	68

There is also an interaction effect for the OPERA question between knowledge and context (sig .001). GENE shows a similar interaction effect (sig .024). However, small N's in both issues for high knowledge mean that this effect is suspect. If we combine high and mid knowledge to increase the N's, for both issues the interaction disappears.

7.2.5 attitude structure effects.

To look at these factors latency was compared using 4-way ANOVAs with questionnaire context, instruction, conflict and importance as the factors.

importance

For ENVIRONMENT there is a main effect of importance (F=8.25, sig. .005) and an interaction between questionnaire context and importance (F=4.99, sig .028).

The main effect of importance is that those who feel the issue is important respond more quickly than those who feel it is unimportant.

Table 7.8
Mean log latency for ENVIRONMENT by importance

	Mean	S.D.	N
Important	2.03	0.45	78
Unimportant	2.42	0.62	21

The interaction between questionnaire context and importance is that those who feel the issue is important take longest to respond in the disagree context while those who think it unimportant take longer to respond in the agree context.

Table 7.9
Mean log latency for ENVIRONMENT
by questionnaire context and importance

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
Important	1.94	0.43	43	2.13	0.46	35
Unimportant	2.54	0.69	7	2.35	0.60	14

Figure 7.1
Mean log latency for Environment
by questionnaire context and importance
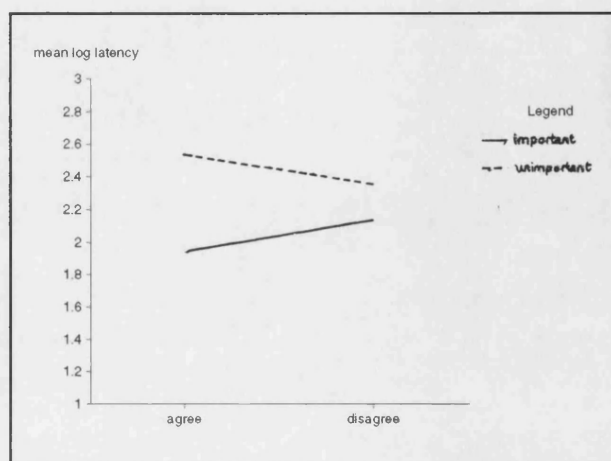
For GENE there is an interaction between questionnaire context
and importance (F=10.92, sig .001). The interaction between
questionnaire context and importance is that those who see the
issue as important respond more quickly in the agree context
than in the disagree context whereas those who see the issue
as unimportant differ little across contexts.

Table 7.10
Mean log latency for GENE
By questionnaire context and importance

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
Important	**1.72**	0.41	29	**2.14**	0.67	25
Unimportant	**1.98**	0.37	21	**1.89**	0.45	25

Figure 7.2

Mean log latency for GENE
By questionnaire context and importance
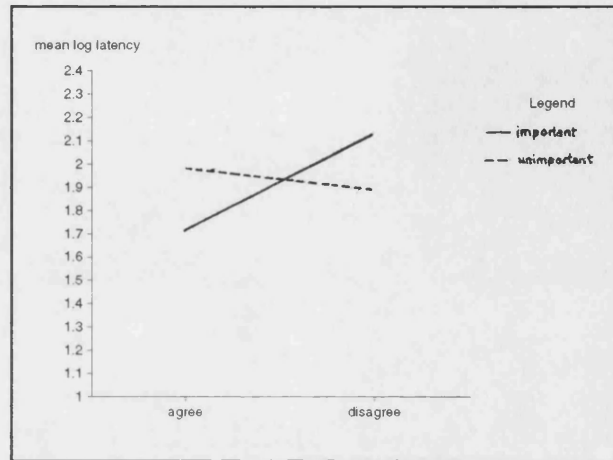
Conflict

For CRIME there is an interaction between instruction and Conflict (F=4.14, sig .045). Those who are one-sided show little effect of instruction. However, for those who are conflicted we get the expected difference in instruction.

Table 7.11
Mean log latency for CRIME by instruction and conflict

	Fast			Slow		
	Mean	SD	N	Mean	SD	N
One-sided	**1.70**	0.47	31	**1.78**	0.46	27
Conflicted	**1.54**	0.46	19	**2.05**	0.51	23

Figure 7.3
Mean log latency for CRIME
by instruction and conflict
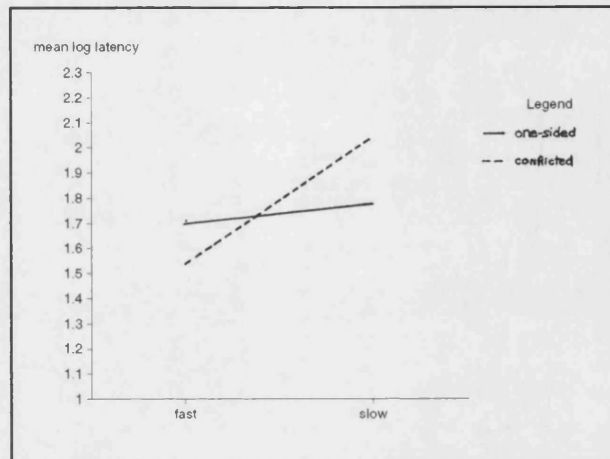
For WHALE there is a main effect of Conflict (F=5.15, sig .026), and a 3-way interaction between questionnaire context, importance and Conflict (F=7.26, sig .009). However, for this interaction, a number of small N's make the finding suspect.

Table 7.12
Mean log latency for WHALE by conflicted

	One-sided	Conflicted
Mean	**2.42**	**2.62**
S.D.	0.44	0.46
N	55	44

The main effect of conflict is that people who are conflicted take longer than those who are one-sided.

For GENE there is an interaction between questionnaire context and Conflict (F=4.06, sig .047). The interaction between questionnaire context and Conflict is that those who are conflicted take more time to respond in the disagree context.

Those who are one-sided take about the same time in each context.

Table 7.13
Mean log latency for GENE
by questionnaire context and conflict

	Agree			Disagree		
	Mean	SD	N	Mean	SD	N
One-sided	**1.78**	0.39	17	**1.82**	0.41	16
Conflicted	**1.86**	0.43	33	**2.11**	0.62	34

Figure 7.4
Mean log latency for GENE
by questionnaire context and conflict
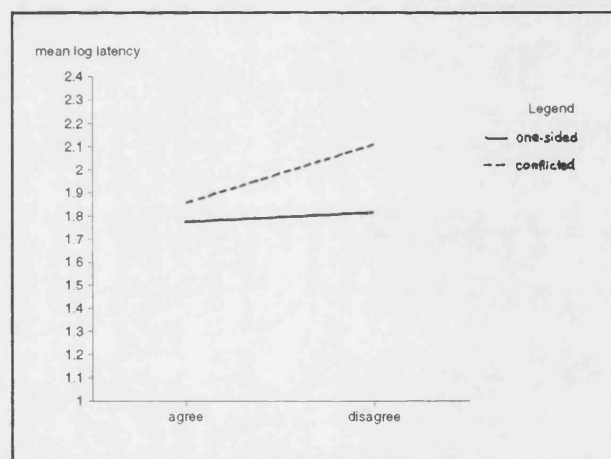
For FOOD there is a main effect of Conflict (F=6.42, sig .013). There is a significant 3-way interaction between questionnaire context, instruction and Conflict (F=6.11, sig .015).

The main effect of Conflict is that those who are conflicted take longer to respond.

Table 7.14
Mean log latency for FOOD by conflict

	One-sided	Conflicted
Mean	**2.06**	**2.24**
S.D.	0.34	0.45
N	56	43

The 3-way interaction between context, instruction and Conflict is that for one sided people there is little difference in the time taken across context although we do get the expected difference across instruction. Conflicted people take more time to respond in the healthy context when given fast instructions and more time to respond in the taste context when given slow instructions.

Table 7.15
Mean log latency for FOOD by questionnaire context, Instruction and conflict

	Healthy			Taste		
	Mean	SD	N	Mean	SD	N
Fast One-sided	**1.95**	0.32	15	**2.00**	0.29	12
Fast Conflicted	**2.20**	0.20	10	**2.08**	0.35	12
Slow One-sided	**2.14**	0.34	15	**2.15**	0.38	14
Slow Conflicted	**2.12**	0.39	10	**2.57**	0.51	11

Summary importance and conflict

There are some, but not a great many, differences in latency from the attitude structure factors of importance and conflict. Where there are effects they are generally in the direction expected. One question, ENVIRONMENT, shows that those who see the issue as important respond more quickly.

For GENE those who see the issue as important respond more quickly in one context whereas those who see the issue as unimportant differ little across context. One would expect important attitudes to be more accessible and thus have a faster latency, ENVIRONMENT may show this. Perhaps important attitudes are more strongly linked to particular aspects of issues, particular contexts, making important attitudes more accessible in some contexts than others. GENE suggests this, where those with important attitudes differ across contexts but those with unimportant attitudes do not.

For conflict, where there is a difference, it is the one-sided people who are quicker (WHALE and FOOD). This fits with the idea that those who are one-sided have only one aspect of the attitude to retrieve and are therefore able to do it more quickly. Where there are interactions with the situational variables (questionnaire context and instruction) the one-sided people show less variation in response time, as would be expected.

Most of the effects of conflict and importance on latency come from media and unfamiliar issues with obvious contexts. So, it may be that these factors influence the effect of attitude structure factors on speed of response. However, these results are also various for the different questions. Some show effects of importance, some of conflict, some show main effects, some interactions. This suggests that the structure of particular attitudes is important. That is, while we may suggest that particular types of issues or particular types of

context are likely to be influential, the particular structure of an issue is also likely to influence latency. The structure of particular attitudes is important in the speed of response (as in shaping the effect of context on response).

Whilst these attitude structure factors show little and varied relationships with latency, attitude extremity is more solid in its effect on latency.

attitude extremity

To examine whether attitude extremity is associated with faster response the response scales given at the target questions were collapsed into three categories, from the mid-point to extreme. To check for interactions with questionnaire context and instruction 3-way anovas were done with questionnaire context, instruction, and extremity as factors. For all the questions only one interaction was found; therefore, the results presented are from one-way anovas with extremity as the factor. A priori comparisons were also done to check for a linear relationship. For all twelve target questions those responding at the extremes had the fastest latencies. Means are given below.

Table 7.16
Mean log latency by attitude extremity

	Extreme	Somewhat	Mid-point	sig level	
Environment	1.88	2.13	2.19	.086	*
Crime	1.60	1.84	1.93	.017	*
EC	1.95	2.05	2.13	.482	
Pornography	1.88	1.93	2.13	.609	
Gene	1.75	1.83	2.03	.003	*
Whale	2.30	2.54	2.72	.004	*
Opera	1.56	1.85	1.73	.010	
Newtech	2.11	2.32	2.22	.095	
House	1.68	1.94	1.97	.029	*
Clean	1.43	1.75	1.94	.000	*
Food	1.91	2.19	2.27	.001	*
Text	1.88	1.99	2.11	.229	

* denotes significant linear relationship

Seven out of the twelve issues show significantly faster
latencies for extreme attitudes. In addition, seven of the
twelve issues show a significant linear relationship, so that
the more extreme the response the faster the latency. This
fits with results showing that extremity is related to
response time, suggesting that extreme attitudes are more
accessible.

7.2.5 Assimilation /Contrast and latency.

The instructions made little difference to response effects.
Only one question (CRIME) showed an effect of instruction.
This effect was such that those given the fast instruction
assimilated whilst those given the slow instruction
contrasted. This finding is in accord with what one would
expect given the supposition that contrasting is a more

effortful process than assimilation. One might expect that more effort is associated with slower latency.

To look at this question of whether contrasting takes longer more generally, we can look at the latencies of those who contrasted with those who assimilated. One simple, if slightly crude, way of defining assimilation and contrast is simply to define those who respond on the positive side of the scale in a positive context as assimilating and so on (strictly speaking this is the definition of assimilation/contrast within the CASM framework). Whilst this measure is certainly not a perfect one it allows us to say something about the amount of time associated with assimilation and contrast.

To compare latencies for assimilation and contrast the scales were split at the midpoint. Those who responded positively in the positive context and those who responded negatively in the negative context were defined as assimilating and vice versa for contrast. 3-way anovas were done with questionnaire context, instruction, and assimilation/contrast as independent factors. The results are presented in table 7.17 below.

Table 7.17
Mean log latency for assimilation and contrast

QUESTION	Assimilate		Contrast			
	Mean	N	Mean	N	F	Sig
Environment	2.13	22	2.03	46	1.85	.178 *
Crime	1.66	38	1.81	41	1.16	.285
Pornography	1.85	44	1.98	37	1.92	.170 *
EC	1.93	35	2.12	36	2.66	.108
Whale	2.55	43	2.33	37	5.27	.025
gene	1.91	29	1.72	39	3.40	.070
opera	1.78	33	1.69	33	.62	.435 *
newtech	2.12	37	2.34	34	5.93	.018
house	1.73	45	2.00	32	5.12	.027
clean	1.73	28	1.60	37	1.55	.218
text	1.90	47	1.93	45	.71	.401
food	2.13	37	2.03	33	.23	.630

*indicates interaction effect

From this measure we can see that for many questions contrasting takes less time than assimilation, however, in only one case (WHALE) is this significantly shorter while another approaches significance (GENE). For some questions contrast takes longer, significantly so in 2 cases (NEWTECH and HOUSE).

Two questions have interactions between assimilation / contrast and instruction. These are ENVIRONMENT and PORNOGRAPHY. For ENVIRONMENT the interaction is such that for those given the fast instruction there is little difference in latency between assimilation and contrast. For those given the slow instruction, assimilation takes longer (F=6.57, sig .013).

Table 7.18
Mean log latency for ENVIRONMENT by instruction
and assimilation/contrast

	Assimilation		Contrast	
Instruction	Mean	N	Mean	N
Fast	1.95	13	1.96	25
Slow	2.40	9	2.11	21

For PORNOGRAPHY assimilation takes longer than contrast with the fast instruction, but with a slow instruction contrast takes longer (F=3.86, p=.053).

Table 7.19
Mean log latency for PORNOGRAPHY by instruction
and assimilation/contrast

	Assimilation		Contrast	
Instruction	Mean	N	Mean	N
Fast	1.80		1.72	
Slow	1.91		2.22	

For the OPERA question there is an interaction with questionnaire context and assimilation/contrast. In the agree context contrasting takes longer than assimilation; in the disagree context assimilation takes longer (F=7.29, sig.009). This may simply reflect that, on this unimportant issue, agreeing is quicker.

Table 7.20
Mean log latency for OPERA by questionnaire context
and assimilation/contrast

	Assimilation		Contrast	
questionnaire	Mean	N	Mean	N
agree	1.64		1.83	
disagree	1.88		1.52	

For the questions where we actually get contrast effects (ENVIRONMENT and GENE) contrast takes a shorter time, though only approaching significance in one case.

7.3 DISCUSSION LATENCY RESULTS

These results provide information about a number of hypotheses concerning response time. Firstly, individual attitude structure seems to play some part in response time. The most consistent relationship is with attitude extremity. As others have shown, more extreme attitudes are answered more quickly. This relationship cuts across knowledge and context types. Importance and conflict have some effect and where they do it is in the direction one would expect. That is, those who see the issue as important respond more quickly. Also, as might be expected, it is those who are one-sided who have faster latencies compared with those who are conflicted. Where differences appear across context or instruction, it is those who are conflicted and those who find the issue important who show differences in response time. It is these people who would also be expected to show more response effect.

Yet, although most effects are for media or familiar issues with obvious contexts, there is little pattern in why particular attitudes or issues show effects of importance and conflict on latency. It may be that these differences are due to the structure of particular issues.

As in looking at response effects across knowledge and context types, the latency results across these factors should be treated as suggestive rather than conclusive in that we are comparing a limited number of different questions. However, they do address some important issues.

It has been suggested that previously accessing a related issue shortens response time because the previous context makes relevant information more accessible. These results suggest that this relation may need to be considered more carefully.

Unfamiliar issues with an obvious context took longest to respond to and for media and familiar issues targets with an obvious context were generally answered no more quickly than those with a subtle context. In comparing context and target latencies, overall, target issues asked after an obvious context were answered no more quickly than the context questions, whereas target questions asked after a subtle context were responded to quicker than the context questions.

The fact that target latencies in subtle contexts are either faster or similar to target latencies in an obvious context,

suggests that an obvious context can induce more thinking about an issue, perhaps an attempt to relate accessed information to the current question. This seems likely to be the case for unfamiliar issues with a more variable situation for media issues, but not the case for familiar issues.

Tourangeau et al (1991) looked at the latencies of response to the issues of abortion and welfare depending on whether the preceding context was either drawn from the same cluster, a different cluster, or was an issue related to or unrelated to the target issue. They found that when the previous items were drawn from the same cluster, latencies were faster. However, they also found that, whilst for abortion latencies were faster following a related issue, for welfare latencies were actually slower in this condition than when the context was unrelated. They were uncertain whether this result was important since it occurred on only one issue. The present findings suggests that this was not a chance finding. Tourangeau et al's study offers a stronger comparison of the effects of different contexts on an issue in that they compared different types of contexts for the same issue, however, they looked only at two, similar, issues. The present results suggest that, as in Tourangeau et al's finding, a related context does not necessarily make response to a related issue quicker. Whilst it may make information more accessible, what this suggests is that, at least for survey questions, accessibility does not necessarily translate into quicker latencies.

One reason for this may be that, as Martin (1986) suggests, processing goals may also be important. Where related information is accessible, people may need to decide whether to use it or not, or perhaps even how to use it, which may take longer than when no related context is given.

The more variable nature of thinking on media issues is emphasised by the larger difference in latency across instructions from this group. It seems that people can think more or less about these issues, if asked to, but, they have more difficulty slowing down for familiar issues and more difficulty speeding up for unfamiliar issues. The exception to this latter may be where the issue is seen as particularly unimportant (as at OPERA).

The idea that contrasting is a more thoughtful and therefore more lengthy process has not received unequivocal support by this research. It may simply be that cognitive effort does not necessarily translate into longer thinking time. For some questions assimilation takes longer and for some contrast takes longer. It is difficult to determine whether the differences are simply due to individual issues or whether they can be linked to other factors. Clearly this issue requires further research, however some speculations are possible.

It may be that for issues with an obvious context contrasting may be faster because the obvious context is more likely to make the respondent aware of the priming incident. However,

this may occur only when the question generally receives a high degree of thought. In other words it may take a certain amount of time generally for contrast to become the first choice response, longer thinking may then be required to assimilate. On the other hand when thinking time is generally fast, assimilation may be the first choice response, with contrast taking longer. Thus, for questions which generally take longer to respond to, comparisons between response latencies for assimilation and contrast would show that contrast is quicker. For questions at which response times are generally quick the same comparison would show assimilation being quicker.

Looking at the differences found here between assimilation and contrast, WHALE and GENE, two questions at which contrast is quicker, have the first and third longest latency (see appendix 13 for mean adjusted latencies). ENVIRONMENT, which shows contrast as quicker in the slow condition has the fourth longest latency overall, and the latency in the slow condition is much slower than in the fast condition (mean =1.57 vs 1.97), in fact, here, it is the third longest latency. HOUSE, on the other hand, where we find that contrasting takes longer comes nearer to the middle of these questions, ranked sixth longest latency. For the other media question with an obvious context, CRIME, contrasting takes longer, but not significantly so, but there is also an effect of instruction on response here such that those given the slow instruction contrast. For this issue the overall latency is very fast,

ranked eighth longest latency (in other words the fourth fastest).

Thus it may well be that contrast takes longer than assimilation, people need to become aware of the biasing nature of that context and decide not to use it in some way. However, when a question receives a good deal of thought there may be a further acceptance step, that is people may recognise the leading nature of the context, but decide that it leads in an acceptable direction. Thus the relation of assimilation/contrast effects to thinking time is not straightforward.

7.4 DISCUSSION RESPONSE AND LATENCY EFFECTS

A further comment on response effects due to instructions is possible having examined response latencies. Only one question, a media issue with obvious context (CRIME), showed variation due to instruction differences. It was suggested that instruction may not have much of an effect. However, looking at the variations in latency across instructions, we might expect to find differences due to instruction at media issues, but less so at familiar and unfamiliar issues. It seems that people may be less able (or perhaps willing) to moderate the amount of thinking time they devote to the latter issues. In this case instruction would have less effect. For media issues, however, differences in instruction were larger. It therefore seems that people are more able to moderate the

amount of time they spend thinking about media issues, and thus, instructions might be expected to have more effect here. The fact that the one instruction effect that was found was for a media issue is in line with this. It is probably also the case that there is likely to be variation due to the particular issue in question. Thus, further study of the effects of instruction on media issues may be fruitful.

A further comment on the direction of the effects found in this study is possible. It may be that, in part, where we find assimilation effects, people are thinking more quickly about the issue (at HOUSE and the fast instruction at CRIME) where we get contrast effects (ENVIRONMENT and GENE and the slow instruction at CRIME) people are generally taking more time thinking about the issue.

CHAPTER 8: DISCUSSION & CONCLUSIONS

Bradburn in 1992 reviewed the progress in understanding context effects in surveys following the application of psychological theory, the so called CASM programme. He argues that both the theoretical framework and the methods used are just a beginning, a move in the direction of explanation. The current research can be seen as a contribution to this effort of developing explanatory concepts. It has examined a number of applications of methods for generating data on survey response processes, and it has explore various theoretical perspectives concerning the response process.

In this concluding section I will review the key findings and draw out the implications of these for survey practice. First I will discuss the results of the studies using think aloud protocols to examine response processes, secondly I will discuss the methods used to elucidate the response process and finally I will discuss the study on context effects for attitude questions. This is followed by a discussion of research at the interface of psychology and survey methodology more generally, suggesting areas where further clarification and research might be useful, and commenting on the implications for survey research.

8.1   DISCUSSING THE RESULTS

8.1.1   Results of protocol analysis and survey response

The application of psychological theory to survey methodology attempts to understand the response process. The focus of

interest is on the cognitive processes involved. The aim is to understand how these processes impact upon the reliability and validity of survey responses; and to understand where bias is likely to occur in survey questioning. In the main, split ballot experiments have been used to examine the response process. While this approach is valuable for some issues, there is a need for a broader methodological base to describe and understand response processes. In particular, qualitative methods which provide more fine grained detail of thought processing could be used.

One way to examine the processes involved in response to a question is to use some process tracing technique to provide data on these processes. Ericsson and Simon (1984) provide a comprehensive review of the use of think-aloud procedures as a method for tracing cognitive processing, and develop an explanation of when and how to use these verbal report techniques. Following this work, the present study used verbal report techniques, think-aloud plus immediate retrospection, to provide data on processing. A range of questions, concerning both behaviours and attitudes were asked in an interview. Respondents thought aloud while answering each question and provided a retrospective account after they had responded to each question.

Two issues were addressed with the protocol data. One is how people construct a response to a question and the other is the impact of the questionnaire on response. The primary focus

was on the latter with the former a secondary focus.

Looking at the processes involved in responding to a single question can provide indicative evidence of the types of biases that are likely to operate. In the present study for example, we asked respondents "Last week, how many hours did you spend in lectures and classes". To respond to this question one might hope that respondents would think back to last week and count the number of hours they spent in classes. However, examining the protocols we found that most respondents use their typical behaviour as an anchor in estimating the actual behaviour over the last week. In this way estimates of their behaviour last week are likely to be biased towards what they think is their typical behaviour. One might test this by obtaining objective measures of the behaviour, attendance records for example, and compare the estimate provided with both actual behaviour last week and an average over a period of time. One would expect the estimates to be closer to the latter than to the former.

However, whilst the above approach may be revealing about processing in general, it does not illuminate how questionnaire features impact upon response. For example, if we had provided pre-coded categories, rather than an open answer format, when asking "Last week, how many hours did you spend watching television", would different response strategies have been used? Would the particular scale provided be influential? To address these types of question

we need to compare response strategies across questionnaire features.

Much of the work in CASM addresses the issue of the impact of questionnaire features on response. Many aspects of the questionnaire including wording, response scales and the context of preceding questions have been shown to influence responses to questions (eg. Schuman & Presser, 1981). In this study we examined a number of these questionnaire aspects with verbal protocols. Three of the questionnaire aspects which we manipulated were drawn directly from research using split-ballot experiments where response effects had been found and hypotheses developed as to the underlying response processes involved in producing these effects. The aim here was to examine whether evidence for the hypothesised processes could be found in the protocols of respondents. In this way the results from split ballot surveys can be explored at the level of processing. The other three questionnaire aspects which we manipulated have received theoretical, and indirect research attention, however the questions were not drawn from previous CASM research. The aim here was to investigate the response processes and suggest likely possibilities for exploration with split ballot surveys.

An example of one hypothesis which we examined with verbal protocols is that of the 'meaning shift' hypothesis (see Chapter 4). Schwarz et al (1988) argue that for a question concerning a vague behaviour, the scales provided for response

are used by respondents to interpret the meaning of a behaviour. Schwarz et al (1988) for example asked people "how often are you annoyed", with either a high or low frequency scale. Respondents given the high frequency scale reported more annoyance than those given the low frequency scale. In this case high frequency implies trivial instances of annoyance and low frequency implies more serious instances of annoyance. Gaskell et al (1992) using more substantive topic areas found less shift. We used questions from Gaskell et al asking " How often are you annoyed with adverts or commercials on television" followed by either high or low frequency scales, to see if this processing -- using the response scale as an indicator of the meaning of the behaviour -- could be observed in the protocols. We found no direct evidence that people used the scales to infer the meaning of a behaviour, however we did find differences in the general strategies used to respond. Thus the influence of scales seems largely to work outside of conscious awareness. This in itself is an important finding. The way these hypotheses had been stated in the past one might have expected people to use the scales in a conscious way. These results suggests that whilst respondents are not consciously aware of the influence of response alternatives, the response alternatives do influence respondents by affecting the way they think about the question. In particular respondents given the low frequency scale thought more about annoying adverts, or the meaning of annoyance than those given the high frequency scale. The strategies revealed by the protocols, offer some suggestions

as to the production of response effects for these vague behavioural questions, and some modification of Schwarz et al's original hypothesis. The strategies used suggest that some scales are more or less likely to 'fit' with respondents prior definitions of ambiguous terms. When the scale does not 'fit' a more detailed consideration of the behaviour is prompted. This may or may not lead to a shift depending upon whether people can generate examples which are consistent with the extremity implied by the scale. For example, those responding to Schwarz et al's question on experience of annoyance may find it fairly easy to generate examples of annoyance which are relatively extreme, as the low frequency scale implies. The term 'annoyed' is fairly vague and open to a range of interpretation. However, with more substantive issues, which are in some ways less vague, more constrained by experience, it may be more difficult to generate examples which are consistent with the extremity implied by the low frequency scale. For example, although one can generate examples of annoying adverts, these may be considered fairly trivial annoyance. It may be fairly difficult to interpret these as serious annoyance, thus less shift would be expected, and indeed Gaskell et al found less shift for these types of behaviours.

By examining the strategies used for response we can better understand why response effects occur. In general for the various questionnaire aspects manipulated (with one exception) we found differences in the strategies used for

response across conditions. Of the hypotheses examined with protocols the example given above was the one where the strategies most clearly indicated why response effects might occur. In the case of other questionnaire aspects manipulated in this study the link between strategies and the production of a response shift was sometimes less clear, though providing useful guidance for further investigation, (for example, in the case of the 'comparison shift' hypothesis, see Chapter 4) and in one case (assimilation/contrast effects for life satisfaction, see Chapter 5) the protocols did not lead to any suggestions as to why response effects might occur. Thus, the detail on cognitive processes provided by the verbal protocols, can be usefully used for examining the way in which questionnaire features may operate, at the level of conscious processing, to produce response effects, but there are limitations to their use. As well as the limitations on the amount and type of information provided, another limitation of the use of verbal protocols has to do more generally with the use of qualitative data.

A weakness of qualitative research lies in its small sample sizes, generalisability is problematic, and when comparing across conditions there is a problem in deciding whether a difference between conditions is important. We cannot be sure that we have captured the full range of response strategies (indeed we are very unlikely to have captured all the minor strategies), nor can we be sure that the differences in strategies between groups are either important or stable.

These weaknesses contrast with the strengths of large scale surveys providing quantitative data -- generalisability is much greater, and comparisons between conditions can be made with statistical precision. However, the weakness of this data is in the detail it provides. While one has reasonably reliable output measures, one has little detail on how these output measures are produced. For example, while one may reliably produce a meaning shift, and hypothesise that this shift is caused by differential interpretation of the question as a result of manipulating the response alternatives, one cannot provide detail of how this operates. There is, so to speak, a danger of putting processes in people's heads.

The fact that at the level of actual responses we found no response effects in our small samples is not surprising. Given the small shifts identified in split ballots we do not have the power to identify these shifts. However, this does not mean that the detail provided by verbal reports cannot elucidate response effects. The qualitative data can provide detail into how observed patterns in survey response, quantitative data, might be produced at a cognitive level. In this case the pattern is one of a difference across groups depending on features of the questionnaire. How this difference might be manifest at a cognitive level was explored. A move back into the field to explore ideas derived from qualitative findings would be useful. It should be regarded as an iterative process, results from one type of study feeding into the design and interpretation of the other.

The strength of the approach taken in this study lies in the exploitation of different research methods to address aspects of the same issue. The issue in this case is the cognitive processes involved in response to questions. Clearly the use of small samples in the laboratory is unlikely to provide the power needed to test for response effects, the appropriate place for this is large sample field surveys. The field survey allows us to examine the question of whether response effects occur due to variations in questionnaire aspects. It might also allow the exploration of hypotheses as to the nature of the processing that might be involved. However, an attempt to understand how the processes operate, at a cognitive level, necessitates detailed information at that level. The strength of qualitative measures lies in the ability to provide a detailed exploration of individual processing. Combined with field survey results this detail can then be used to understand how response effects, which have been shown to occur, could operate at the cognitive level.

Implications

The implications of this study for survey researchers cover both specific and general issues. For example, in general care needs to be taken in interpreting data about usual and mundane behaviours. If these are measures of prime interest survey researchers should work on developing questions which provide better estimates, and drawing on cognitive theory would be useful for this purpose. In terms of the specific

scales of response alternatives provided to assess behavioural questions, survey researchers may be advised to choose a scale that fits with a 'normal' consensual interpretation of the behaviour, especially for questions concerning vague behaviours. On the other hand, if they wanted to provoke a more considered interpretation of the behaviour, they may be advised to use a scale that did not fit. In general, care is also needed in the interpretation of attitude measures; it seems people do not give much consideration to their response.

Thus protocols, and the psychological theory which underlies the interpretation of protocols, can provide some specific advice about how to ask questions. Survey researchers could usefully use protocols when pre-testing questionnaires, but they must be careful in their use, an issue which will be addressed in the next section. But the ability to give specific advice on how to ask questions is as yet limited. Much more understanding is needed of how people answer questions and of the influence of questionnaire features on responses. Indeed, an important question that needs addressing is how we should want people to answer questions.

In terms of further research into CASM, I think some caution is needed in the interpretation of split ballot experiments in terms of the processes operating to produce response effects. In particular care must be taken not to put the independent variable into people's head; that is to presume that the independent variable actually captures the process in

question. At the level of the experiment it is the different questionnaire forms, the independent variable, which is seen as causing the effect. However, at the level of processing, it is the differences in thinking which produce the response effect. The focus in explaining effects in experimental research is the interpretation of the independent variable. However, it is all too easy to see the independent variable as the direct cause of the effect. The more immediate cause of response effects may be the particular strategy used for response.

Although there are limitations on their use, I think much more use could be made of verbal protocols to examine in detail the results found from split ballot experiments, and also to generate ideas for testing with split ballot questionnaires. In this area where the interest is in understanding the processes involved in response, I think both methods are crucial to a clearer understanding of response. However, the use of verbal protocols rests on the assumption that they are valid measures of processing. I want to turn now to address this issue.

8.1.2 Review of methods used

As CASM is concerned with the cognitive processes involved in responding, the use of process tracing techniques to provide data on these processes would be desirable. Within CASM there has been some use of verbal reports of cognitive processes, however, there is little discussion of the issues involved in

using them to elucidate processes in survey responding, very limited assessment of their validity, and little agreement over which method to use. A very few studies have also used latency measures within the survey. In the present studies we looked at the use of both these methods.

verbal reports of cognitive processes

There is much debate within psychology about the usefulness of verbal reports of cognitive processes. In part this stretches back to debates surrounding introspection. More recently the debate has centred around two thesis offering rival claims about the validity of verbal reports. Ericsson and Simon (1984) regard verbal reports as valid data on cognitive processes when they are given during the task or immediately after its completion. In addition the reports should not require subjects to focus on particular aspects of their thoughts. Given these conditions, Ericsson and Simon regard the verbal reports provided as reports of heeded information. Information which is not heeded will not be reported, and thus the reports provided are limited to reports of more conscious thought. Nisbett and Wilson (1977) on the other hand suggest that people are unable to give accurate information on their thought processes. They regard reports of processing as being produced by lay theories of processing rather than as traces of the actual processing. The crux of the debate on validity rests on the issue of whether the reports are traces of processing or whether they are produced by some other process, such as people's ex-post theories of processing. A further

concern in the case of think-aloud protocols is whether thinking aloud interferes with ongoing processing, thereby changing it.

Assessing the validity of verbal reports is a very complex issue (indeed, as is assessing the validity of any psychological measure). One common measure of the validity of think aloud in the area of problem solving is to compare verbalising and non-verbalising groups. If the output, the problem solving, does not change, it is assumed that thinking aloud has not interfered with processing. However, this really only addresses the problem of interference, it does not address the problem of whether the reports could be produced by some other process.

In the present study an assessment was made of the use of verbal reports of cognitive processes within the survey. We tested four different methods for producing verbal reports, all of which were in basic compliance with the principles of valid reported as specified by Ericsson and Simon (1984). The instructions for verbalisation included two different instructions for think aloud only, one combining think-aloud and retrospection, and one retrospection only.

No direct comparison was made between verbalising and non-verbalising groups. This would have been a more powerful test of whether the use of think-aloud in the survey interfered with processing. However, the necessary comparison would have

been with people who were being interviewed (rather than for example a self-completion questionnaire, since this would have confounded the comparison). The cost of data collection, coupled with the low power that would have been achieved from any comparison with the small sample sizes in the verbalising groups, made this a unrealistic option. A comparison of verbalising and non-verbalising groups in a survey would be desirable. However, in the case of survey response, as opposed to problem solving, some caution is needed in comparing output measures. In problem solving, there may be certain processes which must occur for the problem to be solved. If think aloud interferes, the solution may suffer or alternatively be enhanced, thus producing differences in output between verbalising and non-verbalising groups. In the case of survey response, however, the production of a response is more varied. A response may be produced with no thinking at all or with a great deal of thinking; there is no necessary step to producing a response, because, in a sense, any response will do. Unlike problem solving, with survey questions there are no objective criteria for judging the veridicality of a response. Thus even if a comparison between verbalising and non-verbalising groups produced no difference in output, one must be careful in interpreting this as no difference in processing. Different ways of thinking, in the survey, may produce the same response.

The information examined in this thesis to give pointers to the viability of protocol analysis in surveys included

comparisons of the content of verbalisation, the amount and rate of verbalisation, the amount of time taken to produce a response, and the structure of the language used.

This comparison of verbal reports was limited by not having a non-verbalising group for comparison. However, the assessment of the reports does give some indication of the viability of these reports for use in the survey.

Our results suggest that respondents find it relatively easy to verbalise during a survey: the rate of verbalisation compared favourably to other think aloud studies and to normal verbalisation rates (Ericsson & Simon, 1984). The structure of the language was consistent with expectations of language structure for the reporting of currently heeded thoughts. In terms of the content of the reports, there were few differences between the groups. There was some concern with the reports provided by the retrospection only group. Occasionally, these seemed to be reports of how a person might have thought about something rather than how they actually did. Thus while there was no indication that verbal reports are not valid reports of on-going processing, there was some concern with the retrospective only protocols that some 'slippage' might occur. That is, in the midst of reporting what was actually thought, there may be some tendency to speculate on what was thought. This type of report is more likely to be influenced by theories of thought, as Nisbett and Wilson (1977) describe.

Think aloud plus retrospection seemed to provide the most complete reports. In part there may be a slight increase in think aloud in this condition, but more importantly, the chance to provide a retrospective account after thinking aloud allowed respondents to provide further information on very quick responses. For example, a respondent answering a question about how often they used the library last week very quickly responds 'none'. The retrospective report allowed them to report that they immediately thought they were away last week and so did not use the library.

There was a great deal of variability in the length of the protocols provided. Some respondents said a great deal, others said very little. This variability seemed to stem from three sources. First there is variation in the response task, some people are able to give short responses, others, for whatever reason, have more to think about. Second, there may be some variation across people in the ability or motivation to produce verbal reports. Some people find it somewhat easier to think aloud than others, and others are more or less motivated to do so. Finally, there is variation in the motivation to respond to the questionnaire. All of these may interact. Some people approach the questions with interest and genuinely want to provide good answers; others just want to answer the questions and get away; some find it boring; others are more anxious about being questioned. These variations in the motivation to respond have been found to exist in the survey interview (Krosnick, 1991). Thinking

aloud may reflect these positions, but it does not seem, from our study, that it systematically alters them. Thinking aloud and retrospection may be subject to similar influences that exist in the survey interview.

The second study, where we used protocol analysis to examine differences in response strategies across questionnaire features, provides some additional information on the use of think aloud. In this study we found differences in response strategies, but no awareness on the part of the individual as to the questionnaire feature that produced these strategies. That is, as Nisbett and Wilson (1977) argue, people may be unaware of the influences which determine their response; in the case of experiments this is the independent variable. However, as Ericsson and Simon argue, they can report what they are thinking, even if they cannot report why they are thinking it.

Nevertheless, some care is needed in the collection and interpretation of verbal reports of cognitive processes during survey response. First, Ericsson and Simon (1984) suggest that heeded thoughts will be reported. But, this is probably over optimistic. It is likely that not all heeded thoughts will be reported. This is for a number of reasons. People may experience lapses of attention between the thought and its reporting. If many thoughts are occurring at any one time, some of these may not be reported; they may be too quick to bring to attention and label. Also it is likely that some

private information will be censored. Some people may censor more information than others because they are more anxious in reporting. Think aloud is unlikely to be a useful technique for gaining information on sensitive topics.

Secondly, even though most of the thoughts reported may be of what is or was being thought, some subjects may slip into more hypothetical reporting. That is, reporting what they think they must have thought. This seems to be especially likely in retrospective reports, in particular in circumstances when retrospection does not follow think aloud.

Finally, because of the large variability in reports, think-aloud in the survey seems less useful when one is interested in particular pieces of information. This may be the case with the type of information Bishop (1989) is looking for in his use of verbal protocols.

I think that the protocols are best treated as a whole. Individual pieces of information are much more subject to random fluctuations in the ability, or desire to report, and are more variable across a sample not simply because they are or are not thought about, but because they may be thought about or reported at different levels. Looking at the protocol as a whole, and examining general patterns of thought, one is less at the mercy of these variations in reportability.

Thus, verbal reports are probably a reasonably reliable source of data about cognitive processing during the survey interview, though further testing would be desirable. Care is still needed in their use and their use is somewhat limited.

latency measures in surveys

In the final study on context effects for attitude questions, we looked at the use of latencies as information on cognitive processes. The use of latencies within a survey context seems problematic. In their use within cognitive psychology, and mostly as used by Fazio et al (1986; Fazio, 1990), the response situation is very circumscribed. A very simple stimulus is presented, and the response called for is very general. However, in the survey situation, the task is more complex and diverse. For example, Bassili & Fletcher's (1991) use of latencies to examine the strength of attitudes asked, within an interview situation, the question 'Do you think that large companies should have quotas to ensure a fixed percentage of women are hired, or should women get no special treatment?' Based on their responses to this question respondents were then asked either if they would support it 'even if this meant not hiring the best person for the job' or 'even if it meant that women remained economically unequal'. People who shifted their response, or were inconsistent, between the two questions had slower latencies than those who did not change. They interpret this as indicating that those who shifted had less crystallised attitudes. But there is more to this question than simply deciding where one stands.

There are also possibly pressures to respond consistently or to exhibit inconsistency. Perhaps those who did not shift have a strong motivation to appear consistent whereas those who do shift gave more thought to this issue. Those who did not shift may not have crystallised attitudes but simply have a response set. Those who do shift may give more consideration to the issue because they have a more elaborate attitude structure. The latencies cannot discriminate between these explanations, thus using latencies as a simple measure of attitude strength is problematic.

In this thesis we examined response latency. The analysis of latencies in the present study was somewhat restricted by the design; some interesting questions could only be addressed weakly, or indirectly. However, many of the findings indicate a need for caution in the interpretation of latencies in the survey. For example, a simple manipulation of instruction on whether to take one's time in response or to think carefully produced the intended result (for this study) of shorter or longer latencies respectively. However, looked at another way this suggests that latencies measures may be very susceptible to influences other than a simple association between items in memory, for example, influences such as respondent motivation. Similarly, an obvious context (prior questions) in some cases produced longer latencies than a subtle context. This would be contrary to an accessibility account. Here, there may be effort on the part of the respondent to relate the previous context to the present question, and possibly to decide

whether or not to use the previous information. In the case
of survey response, the response is likely to be multi-
determined, including retrieval of elements from memory,
motivation to think, and decisions of whether to use retrieved
information. Both the results of our study, and an
understanding of the nature of the task suggest the use of
latency measures in surveys requires great care both in design
and interpretation because there is much scope for erroneous
interpretation.

8.1.3   Context effects

One major concern for survey researchers is the prevalence and
strength of context effects in the 'field'. There are
differing views on this issue. Some suggest context effects
are common (eg Tourangeau & Rasinski, 1988) others suggest
they are not pervasive, but equally not rare (eg Schuman &
Presser 1981). Studies which suggest context effects are
common tend to be those which manipulate context in
theoretically driven ways. They use largely topical media
issues with a context designed to increase the accessibility
of one or other aspect of the target issue. Studies which
find context effects less common, such as Schuman and
Presser's study, re-examined different versions of a
questionnaire where a large number of questions just happened
to be asked in different contexts. Here, the variation in
context is not theoretically driven but simply a by-product of
the design of the questionnaires. The problem here is there
is little indication of what it is about the different

contexts that leads or does not lead to effects.

One factor which should, theoretically, influence the likelihood of the occurance of context effects is the relatedness of the context to the target. In terms of the episodic relation of context to target there have been a number of studies which examine the effect of scattering context items, or of 'buffering', separating context and target by a block of unrelated issues. One would expect a reduction in context effects when the context and target is separate. Mostly, though not invariably, this is what has been found (Bishop, 1987; Tourangeau et al 1989a; 1989b).

However, as well as the episodic relation one would also expect that the conceptual relation between target and context would affect the likelihood of context effects. One would expect that a strongly related context would produce more effect than a context which is only weakly related. This supposition has not, however, been adequately tested. Tourangeau et al (1989b) used the correlations between items as a measure of relatedness and suggested that those items which were more correlated led to greater context effect. However, correlation between items may not be an adequate measure of relatedness.

In addition research on attitudes would suggest that familiar issues, issues of which people have direct experience, would be less susceptible to context effects than issues which

people have less experience of.

These concerns are about what types of context and targets are most likely to produce effects. These are important issues which have not received much attention in the CASM research programme. Some research within CASM has looked at what type of people are likely to be most effected by context, hence the focus on individual positions on an attitude. The issue of who is most likely to be affected by context effects is also an important one. However, this question is somewhat removed from normal survey practice. Survey researchers are largely interested in main effects of context, since many other measures, such as interest, and knowledge would not be collected in a survey context.

In the present study context effects due to priming were examined. This study explored the issue of the extent of context effects by drawing questions from different levels of knowledge or experience of an issue and by examining different levels of relatedness between target and context. This allows us to look at context effects over a range of target and context types, rather than focusing on question and context combinations where context effects have often been produced. However, rather than a haphazard association between context and target (as in Schuman & Presser's 1981 review), we devised questions and context to differ in theoretically important ways. Thus we can explore the issue of the prevalence of context effects.

We also looked at other factors which may be important to understanding how context operates to influence response. Although theoretically important to context effects, the amount of time spent responding has not been examined. We manipulated thinking time to examine its interaction with context. The influence on context effects of individual variation in knowledge, and attitude structure factors was also examined.

Unlike many studies within the CASM framework, we found few overall, or main effects of context in this study. This is somewhat problematic, since because of the large number of tests conducted some of the results we found may be due to chance. However, there is some theoretical rationale for where we did and did not find context effects (though this latter is also problematic). Most of the effects we found were for obvious contexts (for example, asking about trivial or serious crimes prior to a question concerning sentencing for crime), though not all obvious contexts produced effects. We also found less effects for familiar issues (for example, student housing). This is, generally, as expected. The few interactions we found with thinking time, individual knowledge and attitude structure factors were in the expected directions (Martin, 1986; Tourangeau et al, 1989a, 1989b. See chapters 1 and 6).

Implications

Survey researchers may take some heart from this study. It

suggests that context effects, especially as a main effect are rather more elusive than some CASM research may suggest (although because interactions with personal positions and thinking time factors have been less examined, there may be more interaction effects). It appears that the context must be fairly obviously related to the target question for it to have a effect. So, by and large, one may be able to judge when context is likely to have an effect by judging the relatedness of the context to the target. Although, the results concerning issue familiarity were not so clear cut, it may be that media issues are likely to generate stronger context effects than familiar or unfamiliar issues, the latter are perhaps more likely to interact with personal positions.

Further research exploring context effects might examine the issues of thinking time and the effects of individual knowledge and attitude structure. On the basis of this research, a fruitful area to research the effects of thinking time is likely to be with media issues in an obvious context. For knowledge and attitude structure factors, a concentration on less familiar issues is advised.

In terms of the relatedness of context to target, the split in relatedness used here suggests that subtle contexts are unlikely to produce response effects except, perhaps, in interaction with knowledge and attitude structure for unfamiliar issues. Further exploration of this split for less familiar issues may be useful. However, this was a first

attempt to define relatedness. Rather than looking at a simple split between obvious and subtle, an attempt might also be made to assess different aspects of relatedness. As the relatedness of context and target is likely to be a key factor in context effects, much more work is needed to develop the concept of relatedness and to define what it means. These results suggest that simply looking at correlations between target and context is not a sufficient definition of relatedness.

8.2   DISCUSSING THE SURVEY METHOD

The above results contribute further explanation to a number of aspects of the response process. But they, along with the discussion of survey methods in chapter one, also suggest a number of areas where further exploration may be beneficial to the study of response processes. I would like in this section to broaden the discussion, looking at some of the issues surrounding survey response and looking at where survey measurement stands as a method.

8.2.1   True response versus constructed response

As discussed in chapter one, the previous view of the foundation of survey response generally regarded questioning as a means of getting out a fact or attitude or some other element that was 'in the head'. Response effects were regarded as technical problems. The input of psychological theory has questioned this assumption, showing instead how response can be constructed at the time of questioning and

may not be drawing out any stable elements. Many factors can contribute to this constructive process.

In this study the main focus has been on features of the questionnaire. The think aloud protocols clearly show how processing is affected by changes in the questionnaire, by altering what is thought about, and show how responding involves the construction of a response. The latency study shows how different contexts can effect response and implicates the time taken to think, personal and social factors, in this process. Thus, as in most other CASM research, these results show that responses may be either retrieved or constructed, and that factors within the questionnaire, as well as other types of factors, contribute to this construction.

Several other factors, which may be important in responding have not been studied here. Perhaps most important is the more overarching situational features of the survey. The conversational context of the interview for example, and respondents general representations of what is expected of them in this situation. The think aloud protocols allude to the importance of this in the general nature of the way in which respondents approach questions, that is in a rather conversational manner, but do not address how this might affect response.

I believe that progress in understanding the nature of responding may be aided by developing the conceptions of the

nature of the elements involved in the constructive process. These include conceptions of the relation of context to thinking, the social nature of issues, and ways of thinking about objects, but also issues of respondent motivation and effort as well as individual, and group, differences need to be further addressed. I will now look at these issues.

8.2.2 context

Schuman (1992) broadens the notion of context, which is often treated as the effect of previous questions on later ones, to include not only the influence of one question on another, but also includes the interviewer, the interview setting and even the historical setting. He suggests that for context effects we are likely to find different types of effects with different theoretical explanations. 'Context' may be a term covering several theoretically unrelated ideas. He suggests the importance of bringing different types of context together to examine how they interact in influencing response. There is, I think, the need to address what context is, both in the way Schuman does by looking at different kinds and levels of context, but also by examining more carefully the way that context relates to thought.

At a more general theoretical level, rather than viewing context as a factor, out there, which impinges on the cognitive system, context can be regarded as part of a relationship. Shannon (1990) suggests that context is neither external nor internal but the interface between thought and the world. This is similar to Mead's (1934) conception of the

relational nature of mind. A similar notion is apparent in Palmerino et al's (1984) conception of the attitude as a relation between person and object. A dynamical systems approach (Ostrom, Skowronski, & Nowak, 1994) considers the cognitive system within a temporal dimension; at any moment the contents of thought depend on what they were the previous moment. With each new experience, consciousness is not simply erased and replaced, but rather flows on from what it previously was. Thus, whatever was there previously forms a part of what is there now. In these types of view context is seen as much more integral to thought. In terms of survey responses a serious question in this view is not only why context does effect response but also why it does not.

Concentration on context in terms of a response effect can lead to the view that sometimes people are affected by context and sometimes not. But, context effects should be seen as more general than merely response effects. Krosnick (1992) reports on a study investigating false consensus effects (thinking others share the same view as oneself) due to question order. In this study there was no overall effect of order on the marginal distributions. However, differences were found when the respondents' reported importance of attitudes was taken into account. He suggests that simply looking at the marginal distributions is not sufficient for determining whether variations in the questionnaire have affected response.

Similar results were found in the present study. In the study

on context effects, we found a number of different types of effects. Some of these would not show up as general response effects, because overall the distribution of responses was not altered by the context alone. In a number of cases it was having particular positions on an issue or differing amounts of knowledge, or thinking about it more or less which led to effects. However, since these effects were in different directions for different factors, they would not appear as general response effects. But they are a response effect. Thus, context may have more effect that previously suggested.

Often it is only when an overall effect is found that any attempt is made to relate this effect to other factors -- what caused this overall effect. Thus the focus on the effects of various questionnaire factors on the overall response distributions may have obscured a more pervasive nature of context. The idea that context, which here I am considering as any aspect of the questionnaire, is influential in the way people think about issues is also apparent in the think aloud protocols. For all but one issue we find differences in response strategies at questions. Two of these have actually been shown to lead to response effects, the other three have not been tested in this way, though, if the responses from our small sample transferred to large samples, two of these effects would be unlikely to manifest response effects. Yet, whether they lead to effects in the response distribution is in some ways a moot point, the fact is they have influenced the way people think. This is strong evidence for the interactive nature of thought. It is not a passive process

that draws on static elements and stable processes to respond. Rather thought is dynamically engaged with the environment; the situation influences not only what is thought about (content) but how one goes about thinking (process). One has to question why, if context affects thinking, it does not affect response. One has to explain not only the instances where context does result in a response effect, but also the instances when it does not.

8.2.3 what factors to examine

The social nature of issues

Much of the focus in explaining context effects has been on the accessibility of information. To have an effect context is presumed to need to be both accessible and applicable. This study shows a variety of ways in which accessible context (defined more widely than simply previous questions to include factors such as response scales) may affect both thought and response. Feldman and Lynch's (1988) notion of diagnosticity is comparable to applicability. They suggest that along with information being accessible the diagnosticity is also important but so too is the availability of alternative inputs. But this begs the question what makes information applicable or diagnostic?

Feldman (1992) recognises the importance of the questions of why and when information is diagnostic. He suggests that to answer this question we will need to explore knowledge structures and their degree of flexibility. I agree with this, but I think the concern should not simply be for

individual knowledge structures. In the latency study we found a variety of context effects, some overall effects and some interacting with other factors. It seems that the particular issue and the particular context are important in producing effects. To explain these effects we need to draw on a knowledge of the structure and content of the particular issues investigated. In this study we looked at how the effects varied due to different levels of familiarity and we looked at how related the context was to the target. Both of these factors seem important in explaining context effects. I would argue that these factors represent not an individual structuring of the issue, but rather are tapping the social nature of issues. The use of issue relatedness and familiarity in this study are only two ways of addressing the structure and content of issues, and whilst the results suggest they are important, more detail is needed to examine other aspects of content and structure. Tourangeau et al (1991) address this issue by looking at the structure of attitudes towards welfare and abortion. They show that these issues involve not only a pro-con dimension, but also involve different clusters of beliefs, which do not map directly onto a pro-con dimension. They show that reaction times are quicker to a target when the context is drawn from the same cluster. This structure, again, should not be seen as any one individual's knowledge structure, but is rather an aggregate level structure, which I would argue reflects the social structuring of these issues.

To understand why one context affects responses to an issue

whilst another does not we need to understand the structure and content of issues. Whilst this may seem to suggest that for each issue we have to think anew about context effects, I do not think this is entirely the case. In this study for example, the degree of familiarity of an issue was suggested as important to context effects. One way of addressing the problem is to look for these types of dimensions on which issues can vary, familiarity is likely to be one of them. Perhaps the importance of an issue or the degree of conflict or consensus that exists about it within a society may also be important factors. Research by Boninger, Krosnick and Berent (1995) suggests that the importance of attitudes to an individual results in part from the individual's social identifications.

In one sense examining the structure and content of issues is addressing one of the criticisms often levelled at an information processing approach which is that the 'information' is rarely defined (Graumann, 1988). What is this information that is processed? Massaro and Cohen (1993) reviewing information processing theory, attempt some definition of information. They draw a distinction between information and data; the former being knowledge within the individual and the latter being in the environment. This distinction, however, divorces the individual from the environment, in the same way as does a view of context as an external, rather than a relational factor. Defining information in terms of its social structure and content recognises the individual as part of that structure. Seeing

the individual more firmly as part of this structure, allows one to more clearly define the information which they are dealing with at particular questions.

This position also addresses a criticism levelled at social cognition, namely how is it 'social'. Schneider (1991) suggests one way of doing this is to look at how social variables affect cognitive processes. One way of doing this in the survey is to look at how the conversational situation influences response, and thought; research on the effects of conversational norms has begun to address this issue. Another way is to acknowledge that knowledge structures are, at least in part, cultural productions. Examining how these knowledge structures influence response is to examine the dynamics of these systems. Thus, what one is aiming for in describing the social structure and content of issues, is to, at least in part, define the social and the historical context (Schuman 1992) in which the interview takes place.

Different types of context

But, it is also important to develop a conception of different types of context effect. A number of researchers deal with types of context, with context defined as preceding questions (Schuman & Presser, 1981; Smith, 1992). What these conceptions essentially are trying to capture is the fact that there are different ways of thinking about an object. For example, one might compare it to other, similar objects along the same dimension; this is the specific-specific relation identified by Schuman and Presser (1981), and is very much

338

related to research in social judgement, where the normally found contrast effects are explained as due to the extremity of the anchor provided by the context. Other types of context may include the specific-general relationship, considering the same object in different ways (for example affective or cognitive aspects), or considering an object in terms of related issues (as in the priming paradigm).

The importance of combining both an understanding of the social structure of an issue with an understanding of the type of context may be best illustrated by an example of a well studied context effect for abortion (Schuman & Presser 1981). Support for abortion in the case where the woman does not want any more children (birth control) is less when this question is asked after a question which asks about support for abortion in the case of birth defect. Often this effect is classified as of the specific-general kind. That is, the question dealing with abortion as birth control is seen as a general question. I think this is an incorrect interpretation of this question, rather, I think it represents another reason for abortion and thus these questions are better seen as ratings along the same dimension (specific-specific). The finding of contrast is consistent with other findings which use this type of structure. Birth defects may be seen as an extreme anchor, which perhaps makes the following reason seem rather trivial by comparison. However, other extreme anchors were used prior to the 'birth control' question which failed to find such large effects, if they found effects at all (Schuman, 1992). These included 'pregnancy due to rape',

'poverty of mother', and 'threat to life of mother'. Looking
at the structure of the abortion issue which Tourangeau et al
(1991) identified, it can be seen that these 'reasons' may be
drawing on different belief clusters, with the birth defect
reason as part of a topical cluster which emphasises the
child's welfare in abortion, and the other reasons likely to
fall under a women's rights focus. Thus in this particular
case the birth defect reason may not only be an extreme
anchor, but may involve a categorical difference in the
anchors, which emphasise different aspects of the abortion
argument. Using both an understanding of the type of effect
and of the social structuring of the issue, seems to aid in
the interpretation of context effects.

Thinking effort and processing goals

Martin and Harlow (1992) suggest that simple accessibility of
information does not capture the complexity of response.
Rather the amount of cognitive effort devoted to response is
also important, as are the processing goals.

But the amount of cognitive effort, again, may vary due to the
social structure of issues. Issues may be more or less easy
to devote effort to. As shown in this study, media issues
were more likely to differ in terms of the time taken across
different instructions. It was suggested that for these
issues, as opposed to unfamiliar and, to a lesser extent,
familiar issues, respondents find it easier to moderate the
amount of thinking they devote to the issue. Similarly people
may want to devote more or less attention to some issues.

The idea that processing goals may influence the way in which people think about issues is also important, and needs to be examined more fully. Processing goals may derive not simply from a particular question, but may derive from respondents general approach to, or representation of what an interview is all about. People seem for example, to treat it more as a conversation than an examination. We do not expect, nor usually get, four page structured replies to a question such as 'do you think Britain has benefited from being a member of the EC?'. Yet we might expect this in an examination situation. Similarly, the goals in an interviewing situation may include such things as stating ones views, regardless of whether they are an accurate answer to the question, or trying to answer as one thinks one should answer. Both of these goals seem to underlie some of the responses given in the think aloud study. Understanding the nature of the interactive situation seems to be an important part of understanding how the respondents goals may operate in producing responses, in that it may define more clearly what these goals are. An attempt to integrate the more sociological approaches to the survey (Briggs 1986; Suchman & Jordan, 1992), which look more at the conversational situation might be appropriate here, however, there would be a need to show how the conversational goals impacted upon thought and response. As Schuman (1992) suggests, showing how different factors (questionnaire, interview setting etc) interact is an important aim for understanding survey response.

Individual and group differences

However, it is also important not to neglect individual differences. Although the social structure and content may define issues at one level, within this structure individuals also have their own positions on these issues, as they have their particular place in a social structure. The importance of individual differences in explaining context effects has been demonstrated in this and other studies.

Examining how issues vary across different groups is equally important. Our homogenous sample allowed the definition of some issues as familiar (eg student housing) that would almost certainly not be familiar to a general population sample; some of our unfamiliar issues may be more familiar to different groups.

8.2.4 The use of surveys

This study, and more generally the research area on which it is based, raises serious questions about the use of surveys as a research method. If responses are often constructed, if there is no 'true' response, is the survey a useful measurement instrument?

One suggestion would be to discard measures in traditional survey terms and use more open ended, qualitative, measures which allow respondents more room to qualify and explain their views. There are several problems with this approach though. As well as losing the value of large scale surveys, it is a mistake to see more qualitative methods as necessarily less

problematic. Even qualitative interviews exist in, and create, context, and thus, the issue of the effects of context on thinking and response still exist.

For those who want to measure attitudes and behaviour, some form of survey still seems an attractive method. One way to view the research in CASM is to see it as an attempt to understand effects in order to eliminate them, and for survey researchers this may be the desired goal. But, whilst this research may lead to 'better' questions, it also suggests that in some cases there is no 'best' question. This is especially so in the case of attitude measurement, where any expression of attitude may simply be a temporary construction based on accessible information.

One suggestion, in terms of the context of attitude questions, is to examine how the context in surveys varies from contexts encountered in everyday life. Schuman (1992) suggests that the contextual forces in surveys are likely to be the same as those occurring elsewhere in life. Feldman (1992), on the other hand, suggests that in surveys, questions may bring normally automatic behaviour under conscious control. The concept of 'self-generated validity' (Feldman & Lynch 1988) claims that the questioning may influence the construction of beliefs, a construction which might not otherwise have taken place. This construction may then influence subsequent behaviour.

Thus, there are mixed views as to how much the survey context

varies from everyday experience. But certainly one could suggest, in terms of the particular contexts used, that introducing a context which is unusual, is likely to produce ways of thinking about an issue which would not normally occur, and thus, do not represent a more probable attitude. If this context was not encountered again (and was not recalled by the respondent), that particular attitude may never be produced again. This suggests that a knowledge of the types of contexts in which particular issues arise in everyday life may be important for survey measurement. This again, suggests the need to understand more fully the social nature of the issues one is measuring.

The question of whether it is better for respondents to give a thoughtful, considered, response, or to give a gut reaction response, may rest on a similar comparison to everyday life. Do people normally consider issues or not? Are some issues more likely to be given consideration than others? The fact that respondents have not considered an issue before being asked about it in a survey does not necessarily mean that the response they give is meaningless. It depends on whether the survey context leads to consideration in a way which would not normally be encountered in everyday life. The fact that respondents treat a survey in a conversational way, suggests this may not be very different from other types of activities which respondents engage in.

But, looking at how context varies in everyday life will not eliminate context effects in attitude measurement, in everyday

life, as in the survey, issues may appear in different contexts. If the different contexts have different implications, they may induce different attitudes. If we measure such an attitude in only one context in a survey, we are only getting a partial picture of that attitude. Thus, it may be a better strategy for survey researchers to use contextual variation as a means for understanding attitudes, rather than to fight against context. As well as trying to control response effects, these effects can also be seen as valuable data. Billiet, Waterplas and Loosveldt (1992), looking at attitudes to cohabitation and marriage, show how experimental variation of the context within a survey can provide a deeper understanding of the attitudes being measured.

As Bodenhausen (1992) notes, questionnaire effects are important not just to survey researchers but also as a source of information about the processes involved in responding. That is they are not just methodological problems, but a source of data about thought processes. The widespread effect of context in a variety of areas of psychological research, and the growing interest in the interactional nature of thought, makes them a very important data source.

Jobe, Tourangeau, and Smith (1993) discuss the contributions of survey research to the understanding of memory. They point out that although some of the controls present in laboratory research are lost, researching memory through the survey may be more ecologically valid, allowing the exploration of a

greater range of events and longer time scales. Reveiwing research within the CASM framework they suggest that survey research can make a substantial contribution to the understanding of memory.

Rather than diminish the value of surveys, at least in one way, response effects actually open up exciting possibilities for new ways to use surveys, especially for social scientists. By examining survey response we can examine how the cognitive system interacts with the world, at least in one situation. The study of context effects in other areas has looked at such a relationship. The interesting aspect of the survey is that it questions people about themselves, their behaviour, beliefs, attitudes, and offers a move away from some of the more artificial tasks often studied.

REFERENCES

Abelson, R.P., Loftus, E.F., & Greenwald, A.G., (1992).
Attempts to Improve the Accuracy of Self-Reports of
Voting. In J.M. Tanur (Ed) *Questions about Questions:
Inquiries into the Cognitive Bases of Surveys*. New York:
Russell Sage.

Allport, G.W. (1935). Attitudes. In C. Murchison, (Ed).
Handbook of Social Psychology. Worcester: Clark
University Press.

Alwin, D.F. (1992). Information transmission in the survey
interview: number of response categories and the
reliability of attitude measurement. *Sociological
Methodology, 22*, 83-118.

Alwin, D.F., & Krosnick, J.A. (1985). The Measurement of
Values in Surveys: A Comparison of Ratings and Rankings.
Public Opinion Quarterly, 49, 535-552.

Alwin, D. F., & Krosnick, J. A. (1989). The Reliability of
Attitudinal Survey Data: The Impact of Question and
Respondent Characteristics. Paper presented at annual
conference of the American Association of Public Opinion
Research, St. Petersburg, Florida.

Bainbridge, L. (1985) Inferring from Verbal Reports to Cognitive Processes. In M. Brenner, J. Brown, and D. Canter (eds). *The Research Interview: Uses and Approaches*. London: Academic Press.

Bakan, D. (1954). A reconsideration of the problem of introspection. *Psychological Bulletin, 51*, 105 -118.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. London: Cambridge University Press.

Bassili, J.N. (1993). Response latency versus certainty as indexes of the strength of voting intentions in a CATI survey. *Public Opinion Quarterly, 57*, 54-61.

Bassili, J.N., & Fletcher, J.F. (1991). Response-time measurement in survey research: A method for CATI and a new look at nonattitudes. *Public Opinion Quarterly, 55*, 331-336.

Belson, W.A. (1981). *The design and understanding of survey questions*. Aldershot: Gower.

Bem, D.J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183-200.

Biehal, G., & Chakravarti, D. (1989). The Effects of Concurrent Verbalization on Choice Processing. *Journal of Marketing Research, 26,* Feb, 84-96.

Bickart, B.A. (1993). Carryover and Backfire Effects in Marketing Research. *Journal of Marketing Research, 30,* 52-62.

Billiet, J.B., Waterplas, L., & Loosveldt, G. (1992). Context Effects as Substantive Data in Social Surveys. In N. Schwartz and S. Sudman (Eds), *Context Effects in Social Psychological Research.* New York: Springer-Verlag.

Bishop, G.F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly, 51,* 220-232.

Bishop, G.F. (1989). Think-Aloud responses to survey questions: Illustrations of a new qualitative technique. Paper presented at the annual conference of the American Association for Public Opinion Research, St. Petersburg, Florida.

Bishop, G.F. (1990). Issue involvement and response effects in public opinion surveys. *Public Opinion Quarterly, 54,* 209-218.

Bishop, G.F., Tuchfarber, A.J., & Oldendick, R.W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly, 50,* 240-250.

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14,* 280-288.

Blumenthal, A.L. (1975). A Reappraisal of Wilhelm Wundt. *American Psychologist, 30,* 1081-1087.

Bodenhausen, G.V. (1992). Infomation-Processing Functions of Generic Knowledge Structures and Their Role in Context Effects in Social Judgment. In N. Schwartz and S. Sudman (Eds), *Context Effects in Social Psychological Research.* New York: Springer-Verlag.

Boninger, D.S., Krosnick, J.A. & Berent, M.K. (1995). Origins of Attitude Importance: Self-Interest, Social Identification, and Vaalue Relevance. *Journal of Personality and Social Psychology, 68,* 61-80.

Boring, E.G. (1953). A history of introspection. *Psychological Bulletin, 50,* 3, 169-189.

Bradburn, N.M. (1985). Response Effects. In P.H. Rossi, J.D. Wright and A.B. Anderson (Eds) *Handbook of Survey Research.* London: Academic Press.

Bradburn, N.M. (1992). What Have We Learned? In N. Schwartz and S. Sudman (Eds), *Context Effects in Social Psychological Research*. New York: Springer-Verlag.

Bradburn, N.M., Rips, L.J., & Shevell, S.K. (1987). Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science, 236*, 157-161.

Bradburn, N.M., & Sudman, S. (1979). *Improving Interview-Method and Questionnaire Design*. San Francisco: Jossey-Bass.

Branscombe, N.R. (1988). Conscious and Unconscious Processing of Affective and Cognitive Information. In K. Fiedler & J. Forgas (Eds.) *Affect, Cognition, and Social Behavior*. Toronto: C.J. Hogrefe.

Brenner, M. (1982). Response Effects of "Role-restricted Characteristics of the Interviewer. In W. Dijkstra and J. van der Zouwen (Eds) *Response Behaviour in the Survey-interview*. London: Academic Press.

Briggs, C.L. (1986). *Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research*. Cambridge: Cambridge University Press.

Brown, D.R. (1953). Stimulus-similarity and the anchoring of subjective scales. *American Journal of Psychology, 66*, 199-214.

Burt, C. (1962). The concept of consciousness. *British Journal of Psychology, 53,* 3, 229-242.

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 55,* 50-79.

Byrne R. (1983). Protocol Analysis in Problem Solving. In J.St. B. T. Evans, (ed). *Thinking and Reasoning: Psychological Approaches.* London: Routledge and Kegan Paul.

Cannell, C.F. (1985a). Interviewing in telephone surveys. In T.W. Beed and R.J. Stimson (Eds) *Survey Interviewing: Theory and techniques.* London: George Allen and Unwin.

Cannell, C.F. (1985b). Overview: response bias and interviewer variability in surveys. In T.W. Beed and R.J. Stimson (Eds). *Survey Interviewing: Theory and techniques.* London: George Allen and Unwin.

Cohen, G. (1989). *Memory in the Real World.* London: Lawrence Erlbaum Associates.

Converse, P.E. (1974). Comment: The status of non- attitudes. *American Political Science Review, 68,* 650-660.

Conway, M. & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology, 47,* 738-748.

Cox, E.P. (1980). The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research, 17,* 407-422.

de Groot, A.D. (1966). Perception and Memory versus Thought: Some Old Ideas and Recent Findings. In B. Klienmuntz (Ed) *Problem Solving: Research, Method and Theory.* New York: John Wiley and Sons.

Dijkstra, W., van der Veen, L., & van der Zouwen, J. (1985). A Field Experiment on Interviewer - Respondent Interaction. In M. Brenner, J. Brown and D. Canter (eds). *The Research Interview: Uses and Approaches.* London: Academic Press.

Dippo, C.S. (1989). The use of cognitive laboratory techniques for investigating memory retrieval errors in retrospective surveys. Paper presented at the ISI conference - 47th Session, Paris.

Dovidio, J.F., & Fazio, R.H. (1992). New Technologies for the Direct and Indirect Assessment of Attitudes. In J.M. Tanur (Ed) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys.* New York: Russell Sage.

Ericsson, K.A.. & Crutcher, R.J. (1991). Introspection and Verbal reports on cognitive processes - two approaches to the study of thinking: A response to Howe. *New Ideas in Psychology, 9*, 57-71.

Ericsson, K.A., & Oliver, W.L. (1986). Methodology for laboratory research on thinking: task selection, collection of observations and data analysis. In R. Sternberg and E. Smith (Eds.), *The Psychology of Human Thought*. Cambridge: Cambridge University Press.

Ericsson, K.A., & Simon, H.A. (1980). Verbal Reports as data. *Psychological Review, 87*, 215-251.

Ericsson, K.A., & Simon, H.A. (1984). *Protocol Analysis: verbal reports as data*. Cambridge, Mass.: MIT Press.

Farr, R.M. (1983). Wilhelm Wundt (1832-1920) and the origins of psychology as an experimental and social science. *British Journal of Social Psychology, 22*, 289-301.

Farr, R.M. (1987). The science of mental life: a social psychological perspective. *Bulletin of the British Psychological Society, 40*, 2-17.

Fazio, R.H. (1990). A Practical Guide to the Use of Response Latency in Social Psychological Research. In C. Hendrick & M.S. Clark (eds). *Review of Personality and Social Psychology, 11*, Newbury Park CA.: Sage.

Fazio, R.H., Herr, P.M., & Olney, T.J. (1984). Attitude accessibility following a self-perception process. *Journal of Personality and Social Psychology, 47,* 277-286.

Fazio, R.H., Powell, M.C. & Herr, P.M. (1983). Toward a Process Model of the Attitude-Behavior Relation: Accessing One's Attitude Upon Mere Observation of the Attitude Object. *Journal of Personality and Social Psychology, 44,* 723-735.

Fazio, R.H., Sonbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229-238.

Fazio, R.H. & Williams, C.J. (1986). Attitude accessbility as a moderator of the attitude-perception and attitude-behavior relation: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology, 51,* 505-514.

Fazio, R.H., & Zanna, M.P. (1978). Attitudinal qualities relating to strength of the attitude-behavior relationship. *Journal of Experimental Social Psychology, 14,* 398-408.

Feldman, J.M. (1992). Constructive Processes as a Source of Context Effects in Survey Research: Explorations in Self-

Generated Validity. In N. Schwartz and S. Sudman (Eds),
Context Effects in Social Psychological Research. New
York: Springer-Verlag.

Feldman, J.M., & Lynch, J.G. (1988). Self-Generated Validity
and Other Effects of Measurement on Belief, Attitude,
Intention, and Behaviour. *Journal of Applied Psychology,*
73, 421-435.

Fenigstein, A., Sheier, M.F., & Buss, A.H. (1975). Public and
private self-consciousness: Assessment and theory.
Journal of Consulting and Clinical Psychology, 43, 522-
527.

Fidler, E.J. (1983). The Reliability and Validity of
Concurrent, Retrospective, and Interpretive Verbal
Reports: An Experimental Study. In P. Humphreys, O.
Svenson, and A. Vari (Eds.), *Analysing and aiding*
decision processes. Amsterdam: North-Holland Publishing
Company.

Fischhoff, B. (1991). Value Elicitation: Is There Anything in
There?. *American Psychologist, 46,* 835-847.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude,*
Intention, and Behavior: An introduction to theory and
research. Reading, MA: Addison-Wesly.

Flanagan, S.C. (1982). Measuring value change in advanced industrial societies: A rejoinder to Inglehart. *Comparative Political Studies, 15,* 99-128.

Galambos, J.A., & Black, J.B. (1985). Using Knowledge of Activities to Understand and Answer Questions. In A.C. Graesser and J.B. Black (Eds.) *The Psychology of Questions.* Hillsdale, NJ: Erlbaum.

Gaskell, G.D., O'Muircheartaigh, C.A., & Wright, D.B. (1994). Survey Questions about the Frequency of Vaguely Defined Events: The Effects of Response Alternatives. *Public Opinion Quarterly, 58,* 241-254.

Gaskell, G.D., Wright, D.B., & O'Muircheartaigh, C.A. (1992). The Reliability of Survey Responses: Applying Psychology to Survey Methodology. Paper presented at the BPS London Conference.

Gergen, K.J. (1982). *Toward Transformation in Social Knowledge.* New York: Springer-Verlag.

Gove, W.R. (1982). Systematic Response Bias and Characteristics of the Respondent. In W. Dijkstra and J. van der Zouwen (Eds), *Response Behaviour in the Survey-interview.* London: Academic Press.

Graumann, C.F. (1988). From knowledge to cognition. In D. Bar-Tal & A.W. Kruglanski (Eds), *The social psychology of knowledge*. Cambridge: Cambridge University Press.

Greenberg, J. & Folger, R. (1988). *Controversial Issues in Social Research Methods*. New York: Springer-Verlag.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and semantics 3: Speech acts*. New York: Academic Press.

Hagenaars, J.A., & Heinen, T.G. (1982) Effects of Role-Independent Interviewer Characteristics on Responses. In W. Dijkstra and J. van der Zouwen (Eds), *Response Behaviour in the Survey-interview*. London: Academic Press.

Hamilton, D.L., Devine, P.G., & Ostrom, T.M. (1994). Social Cogntion and Classic Issues in Social Psychology. In P.G. Devine, D.L. Hamilton, & T.M. Ostrom (eds), *Social Cognition: Impact on Social Psychology*. San Diego: Academic Press.

Hamilton, S., & Fagot, B. (1988). Chronic stress and coping skills: A comparison of male and female undergraduates. *Journal of Personality and Social Psychology, 55,* 819-823.

Hastie, R. & Park, B. (1986). The relationship between memory and judgement depends on whether the judgement task is memory-based or on-line. *Psychological Review, 93,* 258-268.

Herr, P.M., Sherman, S.J., & Fazio, R.H. (1983). On the Consequences of Priming: Assimilation and Contrast Effects. *Journal of Experimental Social Psychology, 19,* 323-340.

Hilton, D.J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107,* 65-81.

Higgins, E.T., Rholes, W.S., & Jones, C.R. (1977). Category accessibility and impression formation. *Journal of Personality and Social Psychology, 13,* 141-154.

Hippler, H-J, and Schwarz, N. (1986). Not Forbidding Isn't Allowing: The Cognitive Basis of the Forbid-Allow Asymmetry. *Public Opinion Quarterly, 50,* 87-96.

Hovland, C.I., Harvey, O.J., & Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology, 55,* 244-252.

Howe, R.B.K. (1991). Introspection: A Reassessment. *New Ideas in Psychology, 9,* 25-44.

Hyman, H.H. (1944). Do They Tell the Truth?. *Public Opinion Quarterly, 8,* 557-559.

Hyman, H.H. (1954). The General Problem of Questionnaire Design. In D. Katz, D. Cartwright, S. Eldersveld & A.M. Lee (Eds). *Public Opinion and Propaganda.* New York: Henry Holt & Co..

Hyman, H.H. (1955). *Survey Design and Analysis.* Glencoe, Illinois: The Free Press Publishers.

Inglehart, R. (1977). *The silent revolution: Changing values and political styles among western publics.* Princeton, NJ: Princeton University Press.

Inglehart, R. (1981). Post-Materialism in an Environment of Insecurity. *The American Political Science Review, 75,* 880-900.

Inglehart, R. (1990). *Culture shift in advance industrial societies.* Princeton: Princeton University Press.

Inglehart, R., & Flanagan, S.C. (1987). Controversies: Value change in industrial societies. *American Political Science Review, 81,* 1289-1319.

Jabine, T.B., Straf, M.L., Tanur, J.M., & Tourangeau, R. (Eds), (1984). *Cognitive aspects of survey methodology:*

Building a bridge between disciplines. Washington DC: National Academy Press.

Jobe, J.B. & Mingay, D.J. (1991). Cognition and Survey Measurement: History and Overview. *Applied Cognitive Psychology, 5,* 1-18.

Jobe, J.B., Tourangeau, R., & Smith, A.F. (1993). Contributions of Survey Research to the Understanding of Memory. *Applied Cognitive Psychology, 7,* 567-584.

Jones, L.W. & Koely, L.M. (1993). Multidimensional scaling. In G. Deren & C. Lewis (eds), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues.* Hove: Lawrence Erlbaum Associates.

Joynson, R.B. (1972). The return of mind. *Bulletin of the British Psychological Society, 25,* 1-10.

Judd, C.M., Drake, R.A., Downing, J.W., & Krosnick, J.A. (1991). Some Dynamic Properties of Attitude Structures: Context-Induced Response Facilitation and Polarization. *Journal of Personality and Social Psychology, 60,* 193-202.

Judd, C.M. & Kulik, J.A. (1980). Schematic effects of social attitudes on information processing and recall. *Journal or Personality and Social Psycholgogy, 38,* 569-578.

Kahneman, D., & Miller, D.T. (1986). Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review, 93*, 136-153.

Kallgren, C.A., & Wood, W. (1986). Access to attitude-relevant information in memory as a determinant of attitude-behavior consistency. *Journal of Experimental Social Pscyhology, 22*, 328-338.

Kelman, H.C. (1979). The role of action in attitude change. *Nebraska Symposium on Motivation, 27*, 117-194.

Kendall, P.A., & Lazarsfeld, P.F. (1954). Problems of Survey Analysis. In D. Katz, D. Cartwright, S. Eldersveld & A.M. Lee (Eds). *Public Opinion and Propaganda.* New York: Henry Holt & Co.

Koffka, K. (1924). Introspection and the method of psychology. *British Journal of Psychology, 15*, 149-161.

Kohler, W. (1930). *Gestalt Psychology.* London: G. Bell and Sons Ltd.

Kolers, P., & Roediger, H.L. (1984). Procedures of the mind. *Journal of Verbal Learning and Verbal Behavior, 23*, 425-449.

Krosnick, J. A. (1988). The role of attitude importance in social evaluation: A study of policy preferences,

presidential candidate evaluations, and voting behavior. *Journal of Personality and Social Psychology, 55,* 196-210.

Krosnick, J. A. (1989). Attitude importance and attitude accessibility. *Personality and Social Psychology bulletin, 15,* 297-308.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5,* 213-236.

Krosnick, J.A. (1992). The Impact of Cognitive Sophistication and Attitude Importance on Response-Order and Question-Order Effects. In N. Schwartz and S. Sudman (Eds), *Context Effects in Social Psychological Research.* New York: Springer-Verlag.

Krosnick, J.A., & Abelson, R.P. (1992). The case for measuring attitude strength in surveys. In J.M. Tanur (Ed), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys.* New York: Russell Sage.

Krosnick, J.A., & Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 52,* 201-219.

Krosnick, J.A., & Alwin, D.F. (1988). A test of the form-resistant correlation hypothesis: ratings, rankings, and

the measurement of values. *Public Opinion Quarterly, 52*, 526-538.

Krosnick, J.A., Boninger, D.S., Chuang, Y.C., Berent, M.K. & Carnot, C.G. (1993). Attitude Strength: One Construct or Many Related Constructs? *Journal of Personality and Social Psychology, 65*, 1132-1151.

Krosnick, J.A., & Schuman, H. (1988). Attitude Intensity, Importance and Certainty and Susceptibility to Response Effects. *Journal of Personality and Social Psychology, 54*, 940-952.

Kulpe, O. (1964/1922). The Modern Psychology of Thinking. In J.M. Mandler and G. Mandler (1964) *Thinking: From association to Gestalt.* New York: John Wiley & Sons.

Langer, E.J. (1978). Rethinking the Role of Thought in Social Interaction. In J. Harvery, W. Ickes, & R. Kidd (Eds), *New directions in attribution research, Vol II.* Hillsdale, NJ: Lawrence Erlbaum.

Lazarsfeld, P.F. (1944). The Controversy over Detailed Interviews -- An Offer for Negotiation. *Public Opinion Quarterly, 8*, 38-60.

Lessler, J.T. (1989). Recall strategies for questions from the consumer expenditure survey. Paper presented at the AAAS conference.

Lewin, K. (1939). Field theory and experiment in social psychology. *American Journal of Sociology*, 44, 868-897.

Lieberman, D.A. (1979). Behaviorism and the Mind: A (Limited) Call for a Return to Introspection. *American Psychologist*, 34, 319-333.

Linton, M. (1986). Ways of searching and the content of memory. In D.C. Rubin (Ed), *Autobiographical Memory*. Cambridge: Cambridge University Press.

Linville, P.W. & Carlston, D.E. (1994). Social Cognition of the Self. In P.G. Devine, D.L. Hamilton, & T.M. Ostrom (eds), *Social Cognition: Impact on Social Psychology*. San Diego: Academic Press.

Loftus, E.F. (1984). Protocol analysis of responses to survey recall questions. In T.B. Jabine, M.L. Straf, J.M. Turner, and R. Tourangeau (eds), *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Loftus, E.F., & Fathi, D.C. (1985). Retrieving multiple autobiographical memories. *Social Cognition*, 3, 280-295.

Loftus, E.F., Fienberg S.E., & Tanur, J.M. (1985). Cognitive Psychology Meets the National Survey. *American Psychologist*, 40, 2, 175-180.

Loftus, E.F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? *Memory and Cognition, 11,* 114-120.

Lombardi, W.J., Higgins, E.T., & Bargh, J.A. (1987). The role of consciousness in priming effects on categorization: assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin, 13,* 411-429.

Lyons, W. (1991). Introspection - A two-level or one-level account?: A response to Howe. *New Ideas in Psychology, 9,* 51-55.

Lyons, W. (1986). *The Disappearance of Introspection.* Cambridge, Mass: MIT press.

Mandler, J.M., & Mandler, G. (1964). *Thinking: From Association to Gestalt.* New York: John Wiley & Sons.

Martin, L.L. (1986). Set/reset: Use and misuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51,* 493-504.

Martin, L.L., & Achee, J.W. (1992). Beyond Accessibility: the Role of Processing Objectives in Judgement. In L.L. Martin & A. Tesser (Eds), *The Construction of Social Judgements.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Martin, L.L., & Harlow, T.F. (1992). Basking and Brooding: The Motivating effects of Filter Questions in Surveys. In N. Schwarz & S.Sudman (Eds), *Context Effects in Social and Psychological Research*. New York: Springer-Verlag.

Martin, L.L., Seta, J.J., & Crelia, R. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology, 59*, 27-37.

Martin, S.L. & Klimoski, R.J. (1990). Use of Verbal Protocols to Trace Cognitions Assoicated with Self- and Surpervisor Evaluations of Performance. *Organizational Behavior and Human Decision Processes, 46*, 135-154

Maslow, A.H. (1954). *Motivation and Personality*. New York: Harper.

Massaro, D.W. & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology, 44*, 383-425.

McKellar, P. (1962). The Method of Introspection. In J.M. Scher (Ed), *Theories of the Mind*. New York: Free Press of Glencoe.

Mead, G.H. (1934). *Mind, Self, and Society*. ed. C.W. Morris. Chicago: University of Chicago Press.

Means, B.M. (1989). Memory issues in survey reports: A cognitive interview technique for reducing response errors. Paper presented at the winter meeting of the American Statistical Association. San Diego, CA.

Means, B., Mingay, D.J., Nigam, A., & Zarrow, M. (1988). A cognitive apporach to enhancing health survey reports of medical visits. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds), *Practical aspects of memory: Current research and issues, Vol. 1.* Chichester: John Wiley.

Means, B., & Loftus, E.F. (1991). When Personal History Repeats Itself: Decomposing Memories for Recurring Events. *Applied Cognitive Psychology, 5,* 297-318.

Millar, M.G. & Millar, K.U. (1990). Attitude change as a function of attitude type and argument type. *Journal of Personality and Social Psychology, 59,* 217-218.

Millar, M.G. & Tesser, A. (1989). The effects of affective-cognitive consist and thought on the attitude-behaviour relation. *Journal of Experimental Social Psychology, 25,* 189-202.

Millar, M.G. & Tesser, A. (1992). The Role of Beliefs and Feelings in Guiding Behavior: The Mismatch Model. In L.L. Martin & A. Tesser (Eds) *The Construction of Social Judgements.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, M. (1952). The Waukegan study of Voter Turnout Prediction. *Public Opinion Quarterly, 16,* 381-398.

Mishler, E.G. (1986). *Research Interviewing: Context and Narrative.* Cambridge, Mass: Harvard University Press.

Moscovici, S. (1976). *La Psychanalyse: son image et son public.* 2nd edition. Paris: Presses Universitaires de France.

Mumma, G.H., Draguns, J.G., & Seibel, R. (1993). Reactive effects of concurrent verbalization in person perception tasks. *European Journal of Social Psychology, 23,* 295-311.

Neisser, U. (1986). Nested Structure in Autobiographical Memory. In D.C. Rubin (ed), *Autobiographical Memory.* Cambridge: Cambridge University Press.

Nisbett, R.E., & Bellows, N. (1977). Verbal reports about causal influences on social judgements: private access versus public theories. *Journal of Personality and Social Psychology, 35,* 613-624.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Olson, J.M., & Zanna, M.P. (1993). Attitudes and attitude change. *Annual Review of Psychology, 44*, 117-154.

O'Muircheartaigh, C.A., Gaskell, G.D., & Wright, D.B. (1992). When Response Alternatives affect Survey Results for Vague Behavioral Frequency Questions. Paper presented at the annual conference of the American Association for Public Opinion Research, Florida.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Osgood, C.E., & Tannenbaum, P.H. (1955). The principles of congruity in the prediction of attitude change. *Psychological review, 62*, 42-55.

Ostrom, T.M. (1987). Bipolar survey items: an information processing perspective. In H.J. Hippler, N. Schwartz, and S. Sudman (Eds), *Social information processing and survey methodology.* New York: Springer Verlag.

Ostrom, T.M., Betz, A.L., & Skowronski, J.J. (1992). Cognitive Representations of Bipolar Survey Items. In N. Schwarz & S.Sudman (Eds), *Context Effects in Social and Psychological Research.* New York: Springer Verlag.

Ostrom, T.M., Skowronski, J.J. & Nowak, A. (1994). The Cognitive Foundations of Attitudes: It's a Wonderful Construct. In P.G. Devine, D.L. Hamilton, & T.M. Ostrom (eds), *Social Cognition: Impact on Social Psychology*. San Diego: Academic Press.

Ottati, V.C., Riggle, E.J., Wyer, R.S., Schwarz, N., & Kuklinski, J.K. (1989). Cognitive and Affective Bases of Opinion Survey Responses. *Journal of Personality and Social Psychology, 57,* 404-415.

Palmerino, M., Langer, E., & McGillis, D. (1984). Attitudes and Attitude Change: Mindlessness-Mindfulness Perspective. In J.R. Eiser (ed), *Attitudinal Judgement.* New York: Springer Verlag.

Payne, J.W., Braunstein, M.C., & Carroll, J.S. (1978). Exploring predecisional behaviour: an alternative approach to decision research. *Organizational Behavior and Human Performance, 22,* 17-44.

Payne, S.L. (1951). *The art of asking questions.* Princeton, NJ: Princeton University Press.

Pearson, R.W., Ross, M., & Dawes, R.M. (1992). Personal Recall and the Limits of Retrospective Questions. In J.M. Tanur (Ed), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys.* New York: Russell Sage.

Phillips, D.L., & Clancy, K. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology, 77*, 921-940.

Pratkanis, A.R., & Greenwald, A.G. (1989). A sociocognitive model of attitude structure and function. *Advances in Experimental Social Psychology, 22*, 245-285.

Raden, D. (1985). Stength-related attitude dimensions. *Social Psychological Quarterly, 48*, 312-330.

Radford, J. (1974). Reflections on Introspection. *American Psychologist, 29*, 245-250.

Rasinski, K.A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly, 53*, 388-394.

Rhine, R.J., & Severance, L.J. (1970). Ego-involvement, discrepancy, source credibility, and attitude change. *Journal of Personality and Social Psychology, 16*, 175-190.

Robertson, S.P., Black, J.B., & Lehnert, W.G. (1985). Misleading Question Effects as Evidence for Integrated Question Understanding and Memory Search. In A.C. Graesser and J.B. Black (Eds), *The Psychology of Questions*. Hillsdale, NJ: Erlbaum.

Roese, N.J., & Olson, J.M. (1994). Attitude Importance as a Function of Repeated Attitude Expression. *Journal of Experimental Social Psychology, 30*, 39-51.

Rokeach, M. (1973). *The Nature of Human Values.* New York: Free Press.

Rosenberg, M.J. (1960). A structural theory of attitude dynamics. *Public Opinion Quarterly, 24*, 319-341.

Rosenberg, M.J. (1969). Covert communication in the Psychology experiment. *Psychological Bulletin, 67*, 356-367.

Rosenberg, M.J., & Hovland, C.I. (1960). Cognitive, affective and behavioral components of attitudes. In C.I. Hovland & M.J. Rosenberg (Eds), *Attitude Organization and Change.* New Haven: Yale University Press.

Rosenthal, R., & Rosnow, R.L. (1969). *Artifact in Behavioral Research.* Orlando, FL: Academic Press.

Rubin, D.C. (1986). *Autobiographical Memory.* Cambridge: Cambridge University Press.

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly 5*, 91-92.

Sample, J., & Warland, R. (1973). Attitudes and the prediction of behavior. *Social Forces, 51,* 292-304.

Sanford, F.H. (1951). The Use of a Projective Device in Attitude Surveying. *Public Opinion Quarterly, 14,* 697-709.

Schank, R.C. (1987). Questions and Thought. In P. Morris (ed), *Modelling Cognition.* New York: John Wiley & Sons Ltd.

Schneider, D.J. (1991). Social Cognition. *Annual Review of Psychology, 42,* 527-561.

Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing; I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Schuman, H. (1966). The Random Probe: A Technique for Evaluating the Validity of Closed Questions. *American Sociological Review, 31,* 2, 218-222.

Schuman, H. (1992). Context Effects: State of the Past/State of the Art. In N. Schwartz and S. Sudman (Eds), *Context Effects in Social Psychological Research.* New York: Springer-Verlag.

Schuman, H., Kalton, G., & Ludwig, J. (1983). Context and Contiguity in Survey Questionnaires. *Public Opinion Quarterly*, 47: 112-115.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording and context.* New York: Academic Press.

Schuman, H., & Scott, J. (1989). Response Effects Over Time: Two Experiments. *Sociological Methods and Research, 17,* 398-408.

Schwarz, N. (1990). Assessing Frequency Reports of Mundane Behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. Clark (eds), *Review of Personality and Social Psychology, 11,* 98-119. Beverly Hills: Sage.

Schwarz, N., & Bienias, J. (1990). What Mediates the Impact of Response Alternatives on Frequency Reports of Mundane Behaviors? *Applied Cognitive Psychology, 4,* 61-72.

Schwarz, N, & Bless, H. (1992). Constructing Reality and Its Alternatives : An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgement. In L.L. Martin & A. Tesser (Eds), *The Construction of Social Judgements.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwarz, N., Hippler, H.J., Deutsch, B., & Strack, F. (1985). Response scales: effects of category range on reported behavior and subsequent judgements. *Public Opinion Quarterly, 49*, 388-395.

Schwarz, N., Knauper, B., Hippler, H.J., Noelle-Neumann, I., & Clark, L. (1991). Rating Scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*, 570-582.

Schwarz, N., Munkel, T., & Hippler, H.J. (1990). What determines a 'perspective'? Contrast effects as a function of the dimension tapped by the preceding questions. *European Journal of Social Psychology, 20*, 357-361.

Schwarz, N., & Strack, F. (1988). The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction. Paper presented at AAPOR meeting, Toronto, Canada.

Schwarz, N., Strack, F., & Mai, H-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55*, 3-23.

Schwarz, N., Strack, F., Muller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: further evidence on informative

functions of response alternatives. *Social Cognition, 6,* 107-117.

Schwarz, N., & Wyer, R.S. (1985). Effects of Rank Ordering Stimuli on Magnitude Ratings of These and Other Stimuli. *Journal of Experimental Social Psychology, 21,* 30-46.

Scott, W.A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds), *Handbook of Social Psychology, Vol. 2.* Reading, Mass: Addison-Wesley.

Shannon, B. (1990). What is Context? *Journal for the Theory of Social Behaviour, 20,* 157-166.

Sherif, M., & Hovland, C.I. (1961). *Social Judgement: Assimilation and contrast effects in communication and attitude change.* New Haven: Yale University Press.

Sherman, S.J., Ahlm, K., Berman, L., & Lynn, S.J. (1978). Contrast effects and their relationship to subsequent behavior. *Journal of Experimental Social Psychology, 14,* 340-350.

Slugoski, B.R., Lalljee, M; Lamb, P., & Ginsburg, G.P. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology, 23,* 219-238.

Smead, R.J., Wilcox, J.B., & Wilkes, R.E. (1981). How Valid
Are Product Descriptions and Protocols in Choice
Experiments? *Journal of Consumer Research, 8*, 37-42.

Smith, A.F., Jobe, J.B., & Mingay, D.J. (1991). Retrieval
from memory of dietary information. *Applied Cognitive
Psychology, 5*, 269-296.

Smith E.R., & Miller F.D. (1978). Limits on Perception of
Cognitive Processes: A Reply to Nisbett and Wilson.
Psychological Review, 85, 355-362.

Smith, E.R., & Zarate, M.A. (1992). Exemplar-Based Model of
Social Judgement. *Psychological Review, 99*, 3-21.

Smith, T.W. (1992). Thoughts on the Nature of Context
Effects. In N. Schwartz and S. Sudman (Eds), *Context
Effects in Social and Psychological Research*. New York:
Springer-Verlag.

Smith, T.W. (1987). That which we call welfare by any other
name would smell sweeter: an analysis of the impact of
question wording on response patterns. *Public Opinion
Quarterly, 51*, 75-83.

Strack, F. (1992). 'Order Effects' in Survey Research:
Activation and Information Functions of Preceding
questions. In N.Schwartz and S. Sudman (Eds), *Context

Effects in Social and Psychological Research. New York: Springer-Verlag.

Strack, F., & Martin, L.L. (1987). Thinking, judging and communicating: A process account of context effects in attitude surveys. In H.J. Hippler, N. Schwartz, and S. Sudman (Eds), *Social information processing and survey methodology.* New York: Springer-Verlag.

Strack F., Martin L.L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgements of life satisfaction. *European Journal of Social Psychology, 18,* 429-442.

Strack, F., Schwarz, N., Bless, J., Kubler, A., & Wanke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23,* 53-62.

Strube, G. (1987). Answering survey questions: the role of memory. In H.J. Hippler, N. Schwartz, and S. Sudman (Eds), *Social information processing and survey methodology.* New York: Springer-Verlag.

Suchman, E.A., & Guttman, L. (1947). A Solution to the Problem of Question "Bias". *Public Opinion Quarterly, 11,* 445-455.

Suchman, L., & Jordan, B. (1992). Validity and the Collaborative Construction of Meaning in Face-to-Face Surveys. In J.M. Tanur (Ed), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance, 23,* 86-112.

Svenson, O. (1989). Eliciting and analysing verbal protocols in process studies of judgement and decision making. In H. Montgomery and O. Svenson (eds), *Process and Structure in Human Decision Making.* New York: John Wiley & Sons Ltd.

Sykes, W., & Morton-Williams, J. (1987). Evaluating survey questions. *Journal of Official Statistics, 3,* 191-207.

Taylor, S.E. (1981). The interface of cognitive and social psychology. In J.H. Harvey (Ed), *Cognition, social behavior, and the environment.* Hillsdale: Erlbaum.

Thomas, W.I., & Znanieki, F. (1918). *The Polish peasant in Europe and America (Vol. 1).* Boston: Badger.

Tourangeau, R. (1984). Cognitive Science and Survey Methods. In T.B. Jabine, M.L. Straf, J.M. Tanur, and R. Tourangeau (Eds), *Cognitive aspects of survey methodology: Building*

a bridge between disciplines. Washington DC: National Academy Press.

Tourangeau, R. (1987). Attitude Measurement: A cognitive perspective. In H.J. Hippler, N. Schwartz, and S. Sudman (Eds), *Social information processing and survey methodology.* New York: Springer Verlag.

Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103,* 299-314.

Tourangeau, R., Rasinski, K.A., & Bradburn, N. (1991). Measuring happiness in surveys: a test of the subtraction Hypothesis. *Public Opinion Quarterly, 55,* 255-266.

Tourangeau, R., Rasinski, K.A., Bradburn, N., & D'Andrade, R. (1989a). Carryover effects in attitude surveys. *Public Opinion Quarterly, 53,* 495-524.

Tourangeau, R., Rasinski, K.A., Bradburn, N., & D'Andrade, R. (1989b). Belief accessibility and context effects in attitude measurement. *Journal of Experimental Social Psychology, 25,* 401-421.

Tourangeau, R., Rasinski, K.A., & D'Andrade, R. (1991). Attitude structure and belief accessibility. *Journal of Experimental Social Psychology, 27,* 48-75.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Clarendon Press.

Turner, C.F., & Martin, E. (Eds) (1984). *Surveying subjective phenomena.* New York: Russell Sage Foundation.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 85,* 1124-1131.

Ursic, M.L., & Helgeson, J.G. (1989). Variability in survey questionnaire completion strategies: a protocol analysis. *Journal of the Market Research Society,* 31, 225-240.

Valentine, E.R. (1978). Perchings and flights: Introspection. In A. Burton and J. Radford (Eds), *Thinking in Perspective: Critical essays in the study of thought processes.* London: Methuen & Co Ltd.

Valentine, E.R. (1992). *Conceptual issues in psychology.* London: Routledge.

Van Deth, J.W. (1983). The Persistence of Materialist and Post-Materialist Value Orientations. *European Journal of Political Research,* 11, 63-79.

Watson, J.B. (1930/1967). *Behaviorism.* London: The University of Chicago Press.

Weber, S.J., & Cook, T.D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin, 77,* 273-295.

Wells, A.J. (1991). Optimal presentation orders for the method of paired comparisons. *British Journal of Mathematical and Statistical Psychology, 44,* 181-194.

White, P. (1980). Limitations on Verbal Reports of Internal Events: A Refutation of Nisbett and Wilson and of Bem. *Psychological Review, 87,* 105-112.

White, P. (1985). The awareness issue and memorial influences upon accuracy of verbal report: A re-examination of some data. *Psychological Reports, 57,* 312-314.

Willis, G.B., Royston, P., & Bercini, D. (1991). The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires. *Applied Cognitive Psychology, 5,* 251-267.

Wilson, T.D., & Hodges, S.D. (1992). Attitudes as Temporary Constructions. In L.L. Martin & A. Tesser (Eds), *The Construction of Social Judgements.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wilson, T.D., Kraft, D., & Dunn, D.S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology, 25,* 379-400.

Wilson, T.D., Lisle, D.J., & Kraft, D. (1990). Effects of self-reflection on attitudes and consumer decisions. *Advances in Consumer Research, 17,* 79-85.

Wofford, J.C., & Goodwin, V.L. (1990). Effects of Feedback on Cognitive Processing and Choice of Decision Style. *Journal of Applied Psychology, 75,* 603-612.

Wood, W. (1982). Retrieval of attitude-relevant information from memory: Effects on susceptibility to persuasion and on intrinsic motivation. *Journal of Personality and Social Psychology, 42,* 798-810.

Wright, D.B., Gaskell, G.D., & O'Muircheartaigh, C.A. (1994). How much is "Quite a bit"? Mapping between numeric values and vague quantifiers. *Applied Cognitive Psychology, 8,* 479-496.

Wright, P. (1980). Message-Evoked Thoughts: Persuasion Research Using Thought Verbalizations. *Journal of Consumer Research, 7,* 151-175.

Wright, P., & Rip, P.D. (1980). Product Class Advertising Effects on First-Time Buyers' Decision Strategies. *Journal of Consumer Research, 7,* 176-188.

Wundt, W. (1894/1977). Lectures on Human and Animal Psychology. In the series: *Significant Contributions To the History of Psychology 1750-1920.* D.N. Robinson (Ed), Washington DC: University publications of America.

Wundt, W. (1904). *Principles of physiological psychology.* New York: Macmillan.

Wyer, R.S., & Srull, T.K. (1986). Human Cognition in Its Social Context. *Psychological Review, 93,* 322-359.

Zaller, J., & Feldman, S. (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science, 36,* 579-616.

Zanna, M.P., & Rempel, J.K. (1988). Attitudes: a new look at an old concept. In D. Bar-Tal & A.W. Kruglanski (Eds), *The social psychology of knowledge.* Cambridge: Cambridge University Press.

APPENDIX 1

VERBAL REPORT INSTRUCTIONS - EXPERIMENT 1

THINK ALOUD INSTRUCTION A

(EXPERIMENTS 1 & 2)

In this interview I am interested in what you think about when you answer some questions that I am going to ask you. In order to do this I am going to ask you to THINK ALOUD as you answer the question.

What I mean by think aloud is that I want you to tell me EVERYTHING you are thinking from the time you first hear the question until you give an answer.

I would like you to talk aloud CONSTANTLY from the time I ask the question until you have given your answer to the question.

I don't want you to try to plan out what you say or try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself.

It is most important that you keep talking. If you are silent for any long period of time I will ask you to talk.

THINK ALOUD INSTRUCTION B

In this interview I am interested in what you think about when you answer some questions that I am going to ask you. In order to do this I am going to ask you to THINK ALOUD as you answer the question.

As soon as I ask the question, please start thinking aloud. The best way to do this is to be as spontaneous as possible.

Tell me everything you are thinking as you are thinking it, even details or sidetracks that seem insignificant or embarrassing. If you think aloud spontaneously, you will soon forget that you are speaking at all.

There is no need to explain to me why you are thinking what you are. You don't have to interpret or justify your approach to a question. Just tell me what you are thinking at the moment.

If you are silent for more than a few seconds, I will remind you by saying: Please tell me what you are thinking.

THINK ALOUD PLUS IMMEDIATE RETROSPECTION - INSTRUCTION C - (EXPERIMENTS 1 AND 2)

The think aloud part of this instruction is exactly the same as think aloud instruction A, given above. The following retrospective instructions are also given:

After you have answered the question I then want to see how much you can remember about what you were thinking from the time I asked you the question until you gave the answer. I am interested in what you actually can REMEMBER rather than what you think you must have thought.

If possible I would like you to tell about your memories in the sequence in which they occurred while you were answering the question. Please tell me if you are uncertain about any of your memories.

I don't want you to work on answering the question again, just report all that you can remember thinking about when answering the question.

In this interview I am interested in what think about when you answer some questions that I am going to ask you. In order to do this I am going to ask you, after each question, to report all that you can remember thinking about while answering the question.

I want you to try to remember as much as you can about what you were thinking from the time I asked you the question until you gave the answer. I am interested in what you actually can REMEMBER rather than what you think you must have thought.

If possible I would like you to tell about your memories in the sequence in which they occurred while you were answering the question. Please tell me if you are uncertain about any of your memories.

THE QUESTIONNAIRE - EXPERIMENT 1

First I would like to ask you a few questions about your use of the library.

Q1.   How many hours a week do you usually spend in the library, that is the BLPES?

Q2.   How many hours did you actually spend in the library last week?

Q3.   Would you say that you use the library more than you did last term, less than last term, or about the same?

Q4. In the four weeks ending Sunday (DATE) did you do any of these things: Yes   No

Use the libertas service

Use the card catalogue

Use the microfiche

Ask library staff for help in locating a reference

Ask others for help in locating a reference

Do anything else to locate a reference

Q5.   How often are you annoyed with the services provided by the library?

PROMPT AS NECESSARY -   WOULD YOU SAY YOU WERE ANNOYED:

every day

at least once a week

less than once a week

or never.

Now I'd like to ask you a few questions about your views on some of the issues facing Britain.

Q6. On the whole, to what extent would you say you are

satisfied with the way democracy works in Britain?

Please use the scale on this card to decide on your reply.

'10' means you are completely satisfied and '1' that you

are completely dissatisfied. (SHOW CARD)

completely   completely

dissatisfied 1 2   3   4   5   6   7   8   9   10 satisfied

Q7. Taking everything into consideration, would you say

that Britain has on balance benefited or not from being

a member of the European community?

Q8. If you were to be told tomorrow that the European

Community had been scrapped, would you be very sorry

about it, indifferent, or relieved?

Q9. There is a lot of talk these days about what the goals

of this country should be for the next 10 or 15 years.

On this card are listed some of the goals that different

people say should be given top priority. Would you

please say which one of them you yourself consider most

important in the long run. (SHOW CARD)

 1. Maintaining law and order in the nation

 2. Giving the people more say in important government

 decisions.

 3. Fighting rising prices.

 4. Protecting freedom of speech.

MEAN NUMBER OF WORDS PER QUESTION - EXPERIMENT 1

THINK ALOUD

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
GROUP 1	48.8	35.2	23	38.9	60.9	77.7	85	36.7	111.2
GROUP 2	24.6	16.9	17.3	32.1	66	65	41.1	46.1	101.1
GROUP 3	44.5	22.3	41.2	31.9	97.4	136.4	106.6	22.8	172.7
GROUP 4	9	15.6	11.2	25.8	58.6	24.1	45.8	12.7	10.2

RETROSPECTION

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
GROUP 3	55.3	52.7	42.5	39.3	83.1	96.5	93.7	48.6	136.4
GROUP 4	51.4	48.7	35.4	110.5	90.1	121.0	81.7	53.5	81.7

MEAN NUMBER OF CODES PER QUESTION -EXPERIMENT 1

THINK ALOUD

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
GROUP 1	5.2	5.1	2.6	2.8	6.4	6.8	5.4	3.9	11.1
GROUP 2	2.9	2.3	2.2	2.1	6.9	4.9	4	4	8.7
GROUP 3	5.1	3	3.4	2.7	5.8	8.9	7.7	2.5	14.4
GROUP 4	1.7	2.2	1.4	1.3	3.7	2.8	4.1	1.9	1.4

RETROSPECTION

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
GROUP 3	4.1	3.7	3.2	2.8	4.4	7	6	3.4	8.8
GROUP 4	4.7	4.6	3.4	7.5	5.3	7.6	5.9	3.3	5.2

APPENDIX 4
CODING FRAMES - EXPERIMENT 1

QUESTION 1: USUAL LIBRARY USE

1. Activities
   In library
   general

2. Amount of use
   Hours (response)
   number of visits
   Calculations of amount

3. Personal

4. About the question
   Repeat/rephrase question
   difficulties
   time frame consideration

5. Other

6. Process description

QUESTION 2 -- USE LAST WEEK

1. Activities
   In library
   general

2. Amount of use
   Hours (response)
   number of visits
   Calculations of amount

3. About the question
   Repeat/rephrase question
   difficulties
           time   frame
consideration/
   definition of week

5. Other

6. Process description

QUESTION 3 -- COMPARE USE

1. Activities
   In library
   general

2. Amount of use

3. Personal

4. About the question
   Repeat/rephrase question
   difficulties
   time frame consideration

5. Comments on library

6. Other

7. Process description

8. Response

QUESTION 4 -- FACILITIES USED

1. Activities / Reason for use

2. Amount of use

3. Personal

4. About the question
   Repeat/rephrase question
   difficulties / easy
   time frame consideration

5. Other

6. Process description

7. Responses

QUESTION 5: ANNOYED WITH LIBRARY

1. Aspects / services

2. Defining annoyed

3. Comments on library

4. Personal

5. About the question
   Repeat/rephrase question
     difficulties / easy
       time frame consideration

6. Other

7. Process descriptions

8. Response / Frequency

QUESTION 6:
SATISFACTION WITH DEMOCRACY

1. Aspects of democracy

2. Definitions

3. Comparisons

4. Evaluations

5. Dealing with the scale

6. Personal

7. About the question
   Repeat/rephrase question
     difficulties

8. Other

9. Process Descriptions

10. Responses

QUESTION 7 -- EC BENEFITS

1. Beliefs/aspects considered

2. Considers past or
     future benefits

3. evaluations

4. personal

5. About the question
   Repeat/rephrase question
     difficulties

6. Other

7. Process Descriptions

8. Response

QUESTION 8 -- EC SCRAPPED

1. Beliefs/aspects considered

2. what comes after scrapping

3. evaluations

4. personal

5. About the question
   Repeat/rephrase question
     difficulties

6. Other

7. Process Descriptions

8. Response

394

QUESTION 9 -- GOALS

1. Beliefs/aspects of
   options
      general beliefs
      about priorties

2. evaluations of options

3. Reads / labels options

4. personal

5. About the question
   Repeat/rephrase question
   difficulties

6. Other

7. Process Descriptions

8. Response / responding

QUESTIONNAIRES - EXPERIMENT 2
QUESTIONNAIRE A

1. How many cups or mugs of coffee or tea do you <u>usually</u>
   drink in a day? (SHOW CARD 1A)

                   16 or more cups/mugs a day 1
                   10-15 cups/mugs a day 2
                   4-9 cups/mugs a day 3
                   2-3 cups/mugs a day 4
                   1 cup/mug a day 5
                   Never drink tea or coffee 6

2. Taking your answer from this scale, compared to other
    people, how often would you say that you have coffee or
    tea to drink? (SHOW CARD 2)

    Much more than most 1   A bit less than most ... 4
    A bit more than most.... 2   Much less than most 5
    About average 3

3. Last week, how many hours did you spend watching
   television?   RECORD NUMBER OF HOURS   _____

4. How often do you feel annoyed by an advert or commercial
   on television? (SHOW CARD 4AB1)
        HIGH FREQUENCY   LOW FREQUENCY
   Every day 1   Once a month or more often.. 1
   Most days 2   Once every few months....... 2
   Once a week 3   Once every six months....... 3
   Once a month ... 4   Once a year................. 4
   Less often 5   Less often 5
   Never 6   Never 6

For the next question we simply need to collect some
information about you. For this question, and <u>only</u> this
question, you do not need to think aloud.

5.   Are you .... (READ OUT)..
           Married 1
           In a permanent or long term relationship .. 2
           Or are you single (not in a relationship)? .3

6. Taking things altogether, how would you describe your
   marriage / relationship / dating life. Would you say
   that you very happy, fairly happy, or not too happy with
   your marriage / relationship / dating life?
                                 very happy 1
                                 fairly happy 2
                                 not too happy 3

7. Taken altogether, how would you say things are these
   days? Would you say that you are very happy, fairly
   happy, or not too happy?
                                 very happy 1
                                 fairly happy 2
                                 not too happy 3

Now I'd like to ask you a few questions about your views on some political issues.

8. If you were to be told tomorrow that the European community had been scrapped would you be very sorry about it indifferent or relieved?

                                    very sorry 1
                                    indifferent 2
                                    relieved 3

9. Taking everything into consideration, would you say that Britain has on balance benefited or not from being a member of the European Community?

                                    Benefit 1
                                    Not benefit 2

10. Are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the way democracy works in your country?

                        very satisfied   1
                        fairly satisfied 2
                        not very satisfied 3
                        not at all satisfied .. 4

11. There is a lot of talk these days about what a country's goals should be for the next 10 or 15 years. On this card are listed some of the goals that different people say should be given top priority.
Would you please say which one you yourself consider most important for your country in the long run. (SHOW CARD).
And what would be your second choice.

	First	Second
Maintaining order in the nation	1	1
Giving the people more say in important government decisions	2	2
Fighting rising prices	3	3
Protecting freedom of speech	4	4

QUESTIONNAIRE B.
1. How many cups or mugs of coffee or tea do you <u>usually</u>
   drink in a day? (SHOW CARD)
             4 or more cups/mugs a day 1
             3 cups/mugs a day 2
             2 cups/mugs a day 3
             1 cup/mug a day 4
             Less than 1 cup/mug a day 5
             Never drink tea or coffee 6

2. Taking your answer from this scale, compared to other
    people, how often would you say that you have coffee or
    tea to drink? (SHOW CARD)
    Much more than most 1   A bit less than most ... 4
    A bit more than most ... 2   Much less than most 5
    About average 3

3. Last week, how many hours did you spend in lectures and
      classes?   RECORD HOURS   _____

4. How often do you feel annoyed by an advert or commercial
   on television?   (SHOW CARD)
    HIGH FREQUENCY   LOW FREQUENCY
   Every day 1   Once a month or more often 1
   Most days 2   Once every few months 2
   Once a week 3   Once every six months 3
   Once a month ... 4   Once a year 4
   Less often   5   Less often 5
   Never   6   Never 6

For the next question we simply need to collect some
information about you. For this question, and <u>only</u> this
question, you do not need to think aloud.

5.   Are you .... (READ OUT)..
             Married 1
             In a permanent or long term relationship 2
             Or are you single (not in a relationship)? 3

I would now like to ask you to report on two aspects of your
life, which may be relevant to people's overall well being:
a) marital / relationship / dating satisfaction
b) satisfaction with life as a whole.

6. Taking things altogether, how would you describe your
   marriage / relationship / dating life. Would you say
   that you are very happy, fairly happy, or not too happy
   with your marriage / relationship / dating life?
                               very happy 1
                               fairly happy 2
                               not too happy 3

7. Taken altogether, how would you say things are these
   days? Would you say that you are very happy, fairly
   happy, or not too happy?   very happy 1
                               fairly happy 2
                               not too happy 3

Now I'd like to ask you a few questions about your views on some political issues.

8. Taking everything into consideration, would you say that Britain has on balance benefited or not from being a member of the European Community?

                                        benefit...............1
                                        not benefit.........2

9. If you were to be told tomorrow that the European community had been scrapped would you be very sorry about it indifferent or relieved?

                                        very sorry 1
                                        indifferent 2
                                        relieved.........3

10. On the whole, to what extent would you say you are satisfied with the way democracy works in your country? Please use the scale on this card to indicate your reply. 10 means you are completely satisfied and 1 means you are completely disatisfied. (SHOW CARD)

        completely   completely
        dissatisfied   satisfied
             1 2 3 4 5 6 7 8 9 10

11. There is a lot of talk these days about what a country's goals should be for the next 10 or 15 years. On this card are listed some of the goals that different people say should be given top priority.

    Would you please say which one you yourself consider most important for your country in the long run. (SHOW CARD)
    And what would be your second choice.

	First	Second
maintaining a high level of economic growth ..	1	1
making sure that this country has strong defense forces	2	2
seeing that people have more say about how things are done at their jobs and in their communities	3	3
trying to make our cities and countryside more beautiful	4	4

CONTEXT 1

Environmental issues should be given a higher priority by government.
Strongly Disagree   Strongly Agree
              1   2   3   4 .   5

Motorway and road building poses a great threat to the countryside.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

New jobs should be created even if this sometimes causes damage to the countryside.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

There should be more financial support given to victims and families of victims of drink drivers.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

More rape cases should be tried by women judges.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

Stricter punishment is necessary for many crimes.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

Should the law allow television to show interviews with people who support terrorism in the UK?
   SHOULD allow interviews   1
   SHOULD NOT allow interviews 2

Political parties should be required to disclose their source of income.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

Pornographic magazines and films should be freely available to adults.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

A simpler library system and greater borrowing facilities would help me in my studies.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

The administration at this university is often slow to get things done.
Strongly Disagree   Strongly Agree
              1   2   3   4   5

Do you think Britain has benefited or not from being a member of the EC?

Not Benefited   Benefited
  at all   a lot
     1   2   3   4   5

Research into the development of biological weapons should be banned.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

How likely do you think it is that there will be a serious nuclear accident in the next 10 years?

Very Unlikely   Very Likely
     1   2   3   4   5

Genetic engineering research is adequately controlled.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

Do you believe it's right for western cultures to impose their values onto other non-western cultures?

Right in   Never
all cases   right
     1   2   3   4   5

The west should do more to help african countries where many of the problems stem from past western imperialism.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

The Inuit people of Greenland should be compelled to follow international bans on whaling, rather than be allowed to develop their own rules on whaling.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

Do you agree or disagree with the sale of British antiquities abroad?

Strongly Disagree   Strongly Agree
     1   2   3   4   5

Do you think laws protecting listed buildings should be strengthened?

Should be strengthened   1
Should NOT be strengthened   2

The government should provide more support for the opera.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

The lack of a vision of the future means that most politicians offer only short term policies.

Strongly Disagree   Strongly Agree
     1   2   3   4   5

Many of the problems of modern life stem from people trying
too hard to live in the past.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

Moves to introduce new technologies, such as 'virtual
shopping', into daily life should be resisted.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

The government should do more to house those homeless in
London.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

London's local councils should take more responsibility in
housing the homeless.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

Students in higher education are adequately housed.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

It is right to forbid people living in London from cutting
down trees on their property.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

One of the most attractive features of London is its many
parks and green squares.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

How clean a city do you think London is?
Very Clean   Very Dirty
        1   2   3   4   5

There should be a standard computer operating system so that
computing is made easier for everyone.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

Supermarkets save time and other resources because you can get
most things that you want in one place.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

University courses should be based on a single textbook rather
than lots of different readings.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

Politicians should take decisions for the long term benefit of
people, even it this sometimes means short term suffering.
Strongly Disagree   Strongly Agree
        1   2   3   4   5

Teachers should encourage children to discipline themselves
rather than delivering punishment as this is more likely to
lead to long term self-control.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

On the whole, do you mainly try to eat food which is healthy
or food which tastes good?
Mainly Healthy   Mainly Tastes good
          1   2   3   4   5

CONTEXT 2

Unemployment should be given a higher priority by government.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

Government should finance more projects to create new jobs.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

New jobs should be created even if this sometimes causes
damage to the countryside.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

Extra support should be given to families whose children are
convicted of shoplifting.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

More judges from ethnic minorities should be appointed to try
criminal cases.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

Stricter punishment is necessary for many crimes.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

Do you support or oppose the idea of a law against sex
discrimination in employment, pay and so on?
Strongly oppose   Strongly support
          1   2   3   4   5

Political parties should favour female candidates in order to
balance the representation of women in parliament.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

Pornographic films and magazines should be freely available to
adults.
Strongly Disagree   Strongly Agree
          1   2   3   4   5

The governments of Britain and Ireland should work together more to find a solution to problems in Northern Ireland.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

West Indian immigrants have contributed a lot to British culture.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

Do you think Britain has benefited or not from being a member of the EC?
Not Benefited   Benefited
   at all   a lot
       1   2   3   4   5

Enough money is being provided for cancer research.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

How likely do you think it is that a vaccine will be developed for AIDS in the next 10 years?
Strongly Disagree   Strongly Agree
            1   2   3   4   5

Genetic engineering research is adequately controlled.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

International law - under the UN - should be strengthened.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

Sadaam Hussein should be brought to justice under international law.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

The Inuit people of Greenland should be compelled to follow international bans on whaling, rather than be allowed to develop their own rules about whaling.
Strongly Disagree   Strongly Agree
            1   2   3   4   5

Should the monarchy in Britain be abolished or retained?
      abolished   1
      retained   2

Should the house of lords be abolished or retained?
      abolished   1
      retained   2

The government should provide more support for the opera.
Strongly Disagree   Strongly Agree

```
        1      2      3      4      5
```

The lack of an understanding of history means that most politicians repeat past mistakes.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

Many of the problems of modern life stem from the isolation of individuals from the community.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```


Moves to introduce new technologies, such as 'virtual shopping', into daily life should be resisted.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

The government should lower interest rates further to ease pressure on those paying high mortgages.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

Young people should be given tax relief to assist them in buying their own homes.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

Students in higher education are adequately housed.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

Road pricing schemes should be introduced to reduce traffic congestion in London.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

More pedestrian zones should be created in London to reduce the risk of accidents to pedestrians.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

How clean a city do you think London is?
```
Very Clean                           Very Dirty
          1      2      3      4      5
```

Religious education in schools should cover all the major religions equally.
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

Do you agree or disagree with the saying 'travel broadens the mind'?
```
Strongly  Disagree                   Strongly Agree
          1      2      3      4      5
```

University courses should be based on a single textbook rather than lots of different readings.

Strongly Disagree   Strongly Agree
         1   2   3   4   5

The government should take care of pension provision so that people can get on with living rather than worry about tomorrow.

Strongly Disagree   Strongly Agree
         1   2   3   4   5

It's unfair that wages should rise simply with age because this discriminates against young people who need money to enjoy themselves when they are young.

Strongly Disagree   Strongly Agree
         1   2   3   4   5

On the whole do you mainly try to eat food which is healthy or food which tastes good?

Mainly Healthy   Mainly Tastes good
         1   2   3   4   5

APPENDIX 7

POST-EXPERIMENTAL QUESTIONNAIRE MEASURING KNOWLEDGE,
IMPORTANCE AND CONFLICT.

For this questionnaire the following three questions followed
a description of each issue:

How much do you know about this issue?

1	2	3	4	5
almost	a	something	a lot	a great
nothing	little			deal

For you personally how important is this issue?

1	2	3	4
very	important	not very	not at all
important		important	important

Would you say that your views are mostly on one side or the
other on this issue or would you say that your views are
mixed?

        mostly on one side   1
        mixed   2

The 12 issues which preceded these three questions were:

1. Control of genetic engineering research.

2. The advantages and disadvantages of set texts for
    university courses.

3. Government policy on the availability of pornographic
    material.

4. The benefits to Britain from being a member of the EC.

5. Government support for the opera.

6. Tradeoffs between job creation and damage to the
    countryside.

7. How clean London is.

8. Whaling by the Inuit people of Greenland.

9. Your own choice to eat food because it is healthy or
    because it tastes good.

10. Housing for students in higher education.

11. Criminal justice policy

12. The introduction of new technologies, such as 'virtual shopping' into daily life.

APPENDIX 8

PILOT QUESTIONNAIRE -- EXPERIMENT 3

HOW MUCH DO YOU KNOW ABOUT THESE ISSUES
(Please circle the appropriate response)

Trade offs between job creation and damage to the countryside.

1	2	3	4	5
almost	a	something	a lot	a great
nothing	little			deal

NOTE: the same scale was presented after each issue discription

Control of genetic engineering research.

How clean London is.

The availability of pornographic material.

Britain's benefits from being a member of the EC.

Government support for the opera.

Provision of readings for your courses

Whaling by the Inuit people of Greenland

Your choice to eat food because it is healthy or because it tastes good.

Housing for students in higher education

Punishment for crimes

The introduction of new technologies, such as 'virtual shopping', into daily life.

HOW RELATED ARE THESE ISSUES

a. The risk of serious nuclear accidents.

b. Control of genetic engineering research.

    THESE ISSUES ARE:
         1   2   3   4   5
     not at all   a bit   somewhat   related   very
      related   related   related   related

NOTE: the same scale was presented after each issue pair

- -

a. Creation of pedestrian zones in London.

b. How clean a city London is.

---

a. Justice under inter-national law.

b. The control of whaling by the Inuit people of Greenland
        either by international whaling bans or by the Inuits
        themselves.

---

a. Reducing unemployment.

b. The trade off between job creation and damage to the
        countryside.

---

a. Disclosure of sources of income by political parties.

b. The availability of pornographic material.

---

a. The effect of the government's interest rate policy on
        mortgage payers.

b. Housing for students in higher education.

---

a. The exchange of ideas at university due to the mix of
        nationalities.

b. Britain's benefits from being a member of the EC.

---

a. Laws protecting listed buildings.

b. Government support for the opera.

---

a. Government finance for job creation projects.

b. The trade off between job creation and damage to the

countryside.

---

a. Lecturers encouraging students to be independent.

b. Provision of course reading.

---

a. Politicians repeating past mistakes because of a lack of understanding of history.

b. The introduction of new technologies, such as 'virtual shopping', into daily life.

---

a. Support for victims and families of victims of drink drivers.

b. Punishment for crimes.

---

a. The law banning television interviews with people who support acts of terrorism in the UK.

b. The availability of pornographic material.

---

a. The priority given to environmental issues by government.

b. The trade off between job creation and damage to the countryside.

---

a. The strength of international law -- under the UN.

b. The control of whaling by the Inuit people of Greenland either by international whaling bans or by the Inuits themselves.

---

a. The development of a vaccine for AIDS.

b. Control of genetic engineering research.

---

a. Parks and green squares in London.

b. How clean a city London is.

---

a. The efficiency of the administration at this university.

b. Britain's benefits from being a member of the EC.

---

a. The isolation of individuals from the community as a source of the problems of modern life.

b. The introduction of new technologies, such as 'virtual shopping', into daily life.

---

a. Provision of pensions by the government.

b. Choosing to eat food because it is healthy or because it tastes good.

---

a. Trial of rape cases by women judges.

b. Punishment for crimes.

---

a. The responsibility of London's local councils in housing the homeless.

b. Housing for students in higher education.

---

a. The abolition or retention of the house of lords.

b. Government support for the opera.

---

a. Selection of female candidates by political parties in order to balance the representation of women in parliament.

b. The availability of pornographic material.

---

a. Road pricing schemes to reduce traffic congestion in London.

b. How clean a city London is.

---

a. Wage rises due to age as discriminating against young people.

b. Choosing to eat food because it is healthy or because it tastes good.

---

a. Help from the west for those african countries where many problems stem from past western imperialism.

b. The control of whaling by the Inuit people of Greenland either by international whaling bans or by the Inuits themselves.

---

a. Politicians offering short term policies.

b. The introduction of new technologies, such as 'virtual shopping', into daily life.

---

a. The effects of the library system on your studies.

b. Britain's benefits from being a member of the EC.

---

a. Trial of criminal cases by judges from ethnic minorities.

b. Punishment for crimes.

---

a. The threat to the countryside from motorway and road building.

b. The trade off between job creation and damage to the countryside.

---

a. Financial support for cancer research.

b. Control of genetic engineering research.

---

a. The law in Britain against sex discrimination in employment, pay and so on.

b. The availability of pornographic material.

---

a. Your freedom to express your own views in essays and coursework.

b. Provision of course reading.

---

a. The right of western cultures to impose their values onto other non-western cultures.

b. The control of whaling by the Inuit people of Greenland either by international whaling bans or by the Inuits themselves.

---

a. Discipline of children by encouraging self discipline or by delivering punishment.

b. Choosing to eat food because it is healthy or because it tastes good.

---

a. The sale of British antiquities abroad.

b. Government support for the opera.

---

a. The problems of modern life as a result of people trying too hard to live in the past.

b. The introduction of new technologies, such as 'virtual shopping', into daily life.

---

a. Tax relief to assist young people in buying their own homes.

b. Housing for students in higher education.

---

a. Knowledge of a foreign language as helpful in your studies.

b. Britain's benefits from being a member of the EC.

---

a. Scheduling of classes and lectures.

b. Provision of course reading.

---

a. Support for families whose children are convicted of shoplifting.

b. Punishment for crimes.

- -

a. Control of research into the development of biological weapons.

b. Control of genetic engineering research.

---

a. Government responsibility for housing those homeless in London.

b. Housing for students in higher education.

---

a. The abolition or retention of the monarchy in Britain.

b. Government support for the opera.

---

a. The right of people living in London to cut down trees on their property.

b. How clean a city London is.

---

a. Politicians making decisions for the long term benefit
     of people, even if this sometimes means short term
   suffering.

b. Choosing to eat food because it is healthy or because it
   tastes good.

---

a. The cost of items of stationery from the university
   shop.

b. Provision of course reading.

INSTRUCTIONS FOR COMPLETION OF QUESTIONNAIRE

We are interested in your opinions on a number of different issues. There are no right or wrong answers to these questions. We are interested in your opinions.

FAST INSTRUCTION

There are quite a few questions to get through, so although there is no set time limit, we do ask you to answer each question as quickly as possible. Don't spend time thinking about each issue.

SLOW INSTRUCTION

There are not too many questions and there is no time limit for answering the questions. We ask you to take your time in answering. Think carefully about each issue before you respond.

Press any key to continue

Each question will be presented in a box in the top half of the screen.   Answer categories, or scales, for responding will be presented in a box in the bottom half of the screen.

Select your response from the scale presented by pressing the number key corresponding to your response. This will then be shown in a box within the answer box. You will then be asked to press a key to go on to the next question.

Two practice questions will be given first.

BREAKDOWN OF TARGET QUESTIONS BY IMPORTANCE, CONFLICT, INDIVIDUAL KNOWLEDGE, AND EXTREMITY OF RESPONSE

MEAN IMPORTANCE

1=Very important 4=not at all important

PERCENT WITH MIXED VIEWS

MEAN KNOWLEDGE

MEAN

Bar chart with y-axis labeled MEAN from 1.5 to 4.5. Categories: CLEAN, HOUSE, TEXT, FOOD, CRIME, ENVIRONMENT, EC, PORNOGRAPHY, GENE, WHALE, NEWTECH, OPERA.

1 = Low knowledge   5 = High knowledge

PERCENT GIVING EXTREME RESPONSE

Bar chart with y-axis from 10 to 70. Categories: CLEAN, HOUSE, FOOD, TEXT, CRIME, ENVIRONMENT, EC, PORNOGRAPHY, GENE, WHALE, NEWTECH, OPERA.

418

APPENDIX 11
Interactions of instruction with conflict and
importance for GENE

Table 6.14
Mean response to GENE by
instruction and conflicted

	Fast			Slow		
	Mean	SD	N	Mean	SD	N
one-sided	**2.47**	1.28	17	**1.73**	1.03	15
conflicted	**2.48**	0.97	33	**2.50**	0.93	33

Instructions interact with Conflict (F=6.11, sig .015)

Table 6.15
Mean response to GENE by
instruction and importance

	Fast			Slow		
	Mean	SD	N	Mean	SD	N
important	**2.13**	1.23	24	**2.17**	1.07	29
unimportant	**2.81**	0.80	26	**2.40**	0.94	20

Instructions interact with Importance (F=4.60, sig .035).

CORRELATIONS BETWEEN RESPONSES TO CONTEXT AND TARGET QUESTIONS
-- EXPERIMENT 3

VARIABLE	Questionnare Context 1	Questionnaire Context 2
CLEAN		
context ques1	-.057 (p=.69)	.09 (p=.54)
context ques2	-.016 (p=.91)	.33 (p=.018)
CRIME		
context ques1	.010 (P=.95)	-.378 (P=.007)
context ques2	.043 (P=.77)	-.207 (P= .149)
EC		
context ques1	.013 (P=.928)	.289 (P=.042)
context ques2	.254 (P=.076)	.153 (P=.294)
ENVIRONMENT		
context ques1	-.237 (P=.100)	.272 (P=.059)
context ques2	-.009 (P=.952)	.348 (P=.014)
FOOD		
context ques1	-.055 (P=.708)	.292 (P=.040)
context ques2	.195 (P=.174)	.085 (P=.558)
GENE		
context ques1	.007 (P=.963)	-.028 (P=.851)
context ques2	-.019 (P=.897)	-.159 (P=.276)
HOUSE		
context ques1	-.281 (P=.050)	-.312 (P=.031)
context ques2	-.289 (P=.044)	-.059 (P.687)
OPERA		
context ques1	-.040 (P=.784)	-.049 (P=.738)
context ques2	-.139 (P=.334)	-.036 (P=.810)
PORNOGRAPHY		
context ques1	-.191 (P=.184)	-.088 (P=.543)
context ques2	-.158 (P=.273)	-.051 (P=.725)
TEXT		
context ques1	.327 (P=.020)	-.057 (P=.693)
context ques2	-.021 (P=.885)	-.155 (P=.282)
NEWTECH		
context ques1	.256 (P=.073)	.063 (P=.665)
context ques2	.116 (P=.424)	.323 (P=.023)
WHALE		
context ques1	-.296 (P=.037)	-.040 (P=.784)
context ques2	-.077 (P=.594)	.440 (P=.001)

MEAN ADJUSTED LATENCIES

	FAST	SLOW	OVERALL
OPERA	1.28	1.56	1.42
CLEAN	1.26	1.66	1.46
TEXT	1.37	1.57	1.47
CRIME	1.38	1.71	1.55
PORNOGRAPHY	1.45	1.82	1.64
EC	1.52	1.78	1.65
HOUSE	1.58	1.75	1.67
FOOD	1.57	1.83	1.70
ENVIRONMENT	1.51	1.97	1.74
GENE	1.68	1.84	1.76
NEWTECH	1.81	2.00	1.91
WHALE	1.97	2.13	2.05