What Do People Think about When They Answer Theory of Planned Behaviour Questionnaires?: A 'Think Aloud' Study

David P. French, Richard Cooke, Nikky Mclean, Merroney Williams and Stephen Sutton

*J Health Psychol* 2007; 12: 672
DOI: 10.1177/1359105307078174

The online version of this article can be found at:
[http://hpq.sagepub.com/cgi/content/abstract/12/4/672](http://hpq.sagepub.com/cgi/content/abstract/12/4/672)
What Do People Think about When They Answer Theory of Planned Behaviour Questionnaires?

A ‘Think Aloud’ Study

DAVID P. FRENCH
University of Birmingham, UK
RICHARD COOKE
Aston University, UK
NIKKY MCLEAN
University of Cambridge, UK
MERRONEY WILLIAMS
Sheffield Hallam University, UK
STEPHEN SUTTON
University of Cambridge, UK

Abstract

Two studies aiming to identify the nature and extent of problems that people have when completing theory of planned behaviour (TPB) questionnaires, using a cognitive interviewing approach are reported. Both studies required participants to ‘think aloud’ as they completed TPB questionnaires about: (a) increasing physical activity (six general public participants); and (b) binge drinking (13 students). Most people had no identifiable problems with the majority of questions. However, there were problems common to both studies, relating to information retrieval and to participants answering different questions from those intended by researchers. Questions about normative influence were particularly problematic. The standard procedure for developing TPB questionnaires may systematically produce problematic questions. Suggestions are made for improving this procedure.

Acknowledgements. We are grateful to all participants, and to Paul Norman and Mark Conner for permission to use their questionnaire in Study Two. We are also grateful to Susan Michie and John L. Smith for their insightful refereeing comments, even though they did not necessarily agree with the approach taken or conclusions reported.

Competing interests: None declared.

Address. Correspondence should be directed to:
DAVID P. FRENCH, School of Sport and Exercise Sciences, University of Birmingham, Birmingham, B15 2TT, UK. [Tel. +44 121 414 4128; Fax +44 121 414 4121; email: D.P.French.1@bham.ac.uk]
Introduction

The Theory of reasoned action (TRA; Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) and its extension, the theory of planned behaviour (TPB; Ajzen, 1991; Ajzen & Madden, 1986) are among the most commonly used models for predicting intentions to perform specified behaviours as well as the actual behaviours (see Armitage & Conner, 2001). They have been frequently used to study a variety of health behaviours (Godin & Kok, 1996), and are probably the social cognition models that are most commonly used in health psychology (Johnston, French, Bonetti, & Johnston, 2004; Ogden, 2003).

Part of the appeal of the TPB is that it can be used to predict and explain any behaviour in terms of a limited set of constructs (attitude towards the behaviour, subjective norm, perceived behavioural control and behavioural intention). Another part of the appeal of the TPB is that there is a clearly specified procedure for developing TRA and TPB measures of these constructs for any behaviour, which has been described in detail by Ajzen (1991, 2002) and Ajzen and Fishbein (1980). This procedure involves a phase where those beliefs that are salient in the population of interest are elicited from a pilot sample, with recommended wordings for the questions used in this elicitation. A full questionnaire is then developed, with the beliefs elicited earlier being used to produce questions, alongside more standard questions for the ‘direct’ measures of attitude, subjective norm, perceived behavioural control and intention. Again there are recommendations for the format of questions produced. In contrast to many areas of psychology, tailored TPB questionnaires are created for each new area of application: every time a new behaviour is studied, or the same behaviour is studied with a new population, it is recommended that a new TPB questionnaire is developed (see Ajzen, 2002). Consequently, these questionnaires are rarely, if ever, subject to the same rigorous evaluations of reliability, validity, and sensitivity that are recommended for other psychometric tools. Most TPB questionnaires are used only once or a few times with a specified population and behaviour, which may lead many researchers to view a thorough psychometric evaluation as being unprofitable.

There are undoubtedly strengths, as well as weaknesses, of the standardized approach to developing TPB measures. The use of a standardized procedure makes comparisons of studies using this procedure much simpler, and is very convenient when conducting meta-analyses of these studies. On the other hand, any problems with this procedure do not just affect a single TPB study, but affect all TPB studies in the same way. For example, if this procedure yields common method variance in measures of both attitude and intention and hence inflate estimates of the relationship between attitude and intention, this overestimation of the relationship between attitude and intention will be present in all TPB studies.

The present article reports two studies which share a common objective: to identify what, if any, are the problems that people encounter when they complete TPB questionnaires developed in line with the specified procedure. To achieve this objective, both studies employ cognitive interviewing, where participants are asked to ‘think aloud’ as they complete TPB questionnaires, verbalizing all thoughts that would normally be silent (Ericsson & Simon, 1993; Gilhooly & Green, 1996). Participants are not asked to explain the reasons for their thoughts, or provide any commentary, but just report the information that they are currently thinking about, based on the idea that information currently held in short-term memory can be accurately reported without interfering with the task at hand (Ericsson & Simon, 1993, 1998). Participants’ verbal reports are recorded, transcribed and subsequently coded; in the present case for evidence of misunderstandings of the questions that they are answering.

Protocol analysis of ‘think aloud’ data has been used in a variety of contexts, particularly where participants are attempting to solve the complex reasoning tasks often employed by cognitive psychology. However, it has also been used to assess how people approach the task of completing questionnaires (see Stone et al., 2000). Several models of how people complete questionnaires have been developed from the cognitive psychology literature, and most identify four major component processes: comprehension, retrieval, judgement and response (see Tourangeau, Rips, & Rasinski, 2000). Protocol analysis of ‘think aloud’ data has been used to investigate problems associated with these different processes. For instance, one application of protocol analysis concerned problems with two questionnaires, one concerning radon exposure and one concerning assistive devices, such as wheelchairs or hearing aids (Willis, Royston, & Bercini, 1991). The nature of the problems encountered differed considerably between the two questionnaires. Thus, protocol analysis with the
assistance devices questionnaire revealed mainly problems relating to comprehension, such as inability to understand terminology and variability in interpreting key terms. By contrast, the problems encountered with the radon questionnaire were more linked to retrieval and judgement processes, such as the questionnaire attempting to elicit beliefs about radon that the respondents did not have, and were therefore unable to retrieve.

The aim of the present article is to identify the extent and nature of problems that people encounter when they complete TPB questionnaires that were developed in line with the recommended procedure. In particular, we are trying to identify the nature and extent of any problems that are inherent in the standardized TPB approach. To achieve this aim, two studies are reported, which use cognitive interviewing techniques with participants completing TPB questionnaires, and protocol analysis of the transcripts of verbalizations made by participants. As far as we are aware, this is the first time this approach has been taken with TPB questionnaires. In the two studies, two different questionnaires were used, concerning different behaviours and target populations, and produced by two different research groups, in an attempt to ensure that problems identified are due to the general procedure used to generate TPB questionnaires, rather than the idiosyncrasies of one questionnaire.

The first study employed protocol analysis with a questionnaire developed for an intervention study aiming to increase physical activity (the ProActive trial). All eligible participants had parents with diagnosed type 2 diabetes, and were not very physically active, based on their self-reported levels of activity (see Williams et al., 2004). The intervention was based on changing the beliefs about physical activity that were assessed by the TPB questionnaire, to alter levels of physical activity. The measures contained in the TPB questionnaire were therefore the main process or mediating variables by which the intervention was hypothesized to have its effect on levels of physical activity.

**Study One**

**Methods**

**Participants** The participants were six volunteers from the same population as people participating in an intervention study aiming to increase physical activity. All participants were aged between 30–50 years, and lived in Cambridgeshire, Essex or West Suffolk; two were male, four were female.

**Materials** Participants completed a 42-item TPB questionnaire concerning increasing physical activity (questionnaire available from authors). This questionnaire assessed all 10 standard constructs of the TPB, including belief-based measures (see Table 2), based on a belief elicitation study involving 213 participants (French et al., 2005; Sutton et al., 2003). The questionnaire was designed in accordance with the recommendations of Ajzen (1991) and Ajzen and Fishbein (1980), with the exception of all items being phrased as statements, with responses required in terms of the following five options: ‘strongly agree’; ‘agree’; ‘neither agree nor disagree’; ‘disagree’; and ‘strongly agree’. This deviation from the standard recommendations was made to facilitate ease of responding with a general public population. A definition of physical activity was given at the beginning of the questionnaire.

**Procedure** All participants agreed to be audio-taped while ‘thinking aloud’ and completing the TPB questionnaire. Before beginning, they were read the following instructions, which were adapted from Green and Gilhooly (1996):

We will shortly be beginning a study to persuade people to be more active. The aim of this study is to prevent them from getting diabetes. For this study, we have developed a questionnaire about people’s beliefs about diabetes, which we are trying to change. We want to check that people understand the questions in the way that we meant them. To do this, I am going to ask you to think aloud as you complete the questionnaire. What I mean by ‘think aloud’ is that I want you to tell me everything you are thinking as you read each question and decide how to answer it. I would like you to talk aloud constantly. I don’t want you 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. If you are silent for any long period of time, I will ask you to talk. Please try to speak as clearly as possible, as I shall be recording you as you speak. Do you understand what I want you to do?

Any queries were dealt with at this stage by the researcher, who then sat out of the line of sight of participants, to minimize influence. Once participants...
began completing the questionnaire, they were not interrupted, unless they fell silent for about 10 seconds, in which case they were instructed to ‘keep talking’.

Results

Segmenting and coding of think aloud transcripts

The first two authors independently went through the six transcripts, to segment them into material relating to each of the 42 TPB questions. Overall, there were four disagreements over segmenting, which were easily resolved following discussion. Given that there were \(6 \times 42 = 252\) resulting segments, agreement in segmenting the text was over 98 per cent (i.e. \(100 \times 248/252\)).

Based on an examination of the transcripts of two participants, a coding scheme (shown in Table 1) with four categories was developed for the segments of text. Reasonable agreement with this scheme was achieved between the first two authors, yielding a kappa based on the responses of all six participants (corrected for chance agreement) of 0.62. This level of agreement was arrived at after several iterations of independently coding transcripts, discussion of differences, then re-coding of transcripts. Between four and 18 problematic segments per participant were identified, or put another way, for each participant, there were no problems identified for between 24 and 38 of the 42 questions answered. As Table 1 shows, the three types of problems occurred with approximately equal frequency. The frequency of problematic segments for each of the 10 TPB constructs is shown in Table 2. Given the small number of participants, no further quantitative analyses are reported.

Examples of the problems identified

In the examples that follow, we transcribe the whole segment that was coded. In most cases, this involves the participants first reading aloud the question they were answering.

Participant re-read question, or seriously floundered in answering it

The simplest kind of problems identified involved the participants needing to re-read the question to extract the appropriate meaning, or floundering when trying to answer it. Some problems of this nature arose because of the cognitive complexity involved in comprehension of some items. For instance, some participants found it difficult when they wanted to disagree with a question that was already negatively phrased:

Q41. (subjective norm). Most people whose views I value would disapprove if I was more physically active in the next 12 months. Well they wouldn’t disapprove so no, I disagree. Hang on. Do I disagree or am I agreeing. Hang on. Most people whose views I value would disapprove if I was more physically active in the next … I’d agree. Yes. They would … No. They wouldn’t disapprove. Sorry. I’ve got myself confused now! You’re right. They would disapprove. No. They wouldn’t! God. What have I done? (participant 5, female)

On other occasions, the complexity arose apparently because the belief items contained clauses proposing hypothetical states of affairs:

Q20. (control belief). If I had a lack of interest in being more physically active in the next 12 months it would make it difficult for me to be more. Oh my life. If I had a lack of interest in being more physically active in the next 12 months it would make it difficult for me to be more physically active. Yes. (participant 4, female)

A final type of floundering when answering questions arose where the participants thought their answers depended on further information, which was not specified in the question, such as:

Q39. (behavioural belief). If I was more physically active in the next 12 months it is likely that I would spend more time with my family. Well not … Don’t … Neither agree nor disagree. Just depends on if we choose to do stuff together. (participant 5, female)

Q27. (normative belief). My partner would want me to be more physically active in the next 12 months. Again I’m not really sure because I think it would depend what I wanted to do and how much it encroached in what he wanted to do and our time together. (participant 2, female)

Participant questioned how sensible the questions were

On other occasions, when participants thought their answers depended on further information which was not specified in the question, instead of floundering, they questioned how sensible some of the questions were, or identified problems with how the questions were worded:

Q40 (outcome evaluation). Spending more time with my family would be a good thing. In what...
Table 1. Frequency and type of problematic segments identified for the six respondents in Study One

<table>
<thead>
<tr>
<th>Type of problem</th>
<th>Respondent number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No significant problems identified</td>
<td></td>
<td>33</td>
<td>34</td>
<td>32</td>
<td>38</td>
<td>33</td>
<td>24</td>
<td>194</td>
</tr>
<tr>
<td>Respondent re-read question, or seriously floundered in answering it</td>
<td></td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Respondent identified in self problems in understanding question</td>
<td></td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td>Questioned sensibleness of question</td>
<td></td>
<td>39</td>
<td>33</td>
<td>32</td>
<td>33</td>
<td>32</td>
<td>24</td>
<td>194</td>
</tr>
<tr>
<td>Answered a different question from the one that was asked, or gave reasoning inconsistent with the answer given</td>
<td></td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Respondent had problems in comprehending/answering question, but were not themselves necessarily aware of this</td>
<td></td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td>Any of above problems</td>
<td></td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>4</td>
<td>9</td>
<td>18</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 2. Frequency and type of problematic segments identified in Study One, according to the TPB constructs assessed

<table>
<thead>
<tr>
<th>Construct</th>
<th>Number of q’aire items</th>
<th>Number of problems</th>
<th>Re-read question or floundered</th>
<th>Questioned how sensible question was</th>
<th>Answered different question or question and answer inconsistent</th>
<th>Problems/item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief-based measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioural beliefs</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1.13</td>
</tr>
<tr>
<td>Outcome evaluations</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Normative beliefs</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Motivation to comply</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Control beliefs</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.25</td>
</tr>
<tr>
<td>Control power</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Direct measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Subjective norm</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>Perc. behaviour control</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Intention</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.5</td>
</tr>
</tbody>
</table>

context? Do you mean in the. If that follows on directly from question 39 er well no because members of my family wouldn’t actually want to do the activities that I do and have them around while I was doing those would actually detract from the enjoyment and pleasure. And yet if you’re answering the straight question spending more time with my family would be a good thing without reference to question 39 well then that’s another matter. In a close-knit family you enjoy their company but not all the time so that’s a very badly framed question and you’re not sure what that’s linked to. So neither
agree or disagree on that one simply because of the framing of the question. (participant 6, male)

In this case, the participant is unsure whether the question about spending time with his family refers to the general case, or the specific case of spending time with his family while being physically active. In the next example, the participant points out that ‘spare time’ can be construed as neither, either or both time spent being physically active, and time spent with his children:

Q4. (control belief). If I did not have much spare time in the next 12 months it would make it difficult for me to be more physically active. Well [pause] answering the first three, I don’t know how to answer that one because by doing as much exercise as I do there is more spare time if you spend with the kids … I should think. So I don’t know what to actually answer to that one. [pause] Is it alright to leave that one blank? (participant 1, male)

A further problem occurred when the participants indicated that there was only one coherent way to answer a question, as they thought that their previous answers implied an answer to a subsequent question:

Q12. (outcome evaluation). To improve my health would be good. Well I’d agree with that one as well but that’s… That question is virtually answered by the last three questions itself I mean. (participant 1, male)

Participant answered a different question from the one that was asked, or gave reasoning inconsistent with the answer given

The final kinds of problems identified concerned participants answering a different question from the one that was asked, by using information that the question did not seek. This was particularly apparent with the questions concerning social influence, where people answered questions about what other people wanted them to do by referring to whether this mattered to them:

Q6. (normative belief). My friends would want me to be more physically active in the next 12 months. I think I’ve got to neither agree or disagree with that one because I have to consider it’s none of their business! [laughs] (participant 3, female)

Q34. (subjective norm). Most people who are important to me would want me to be more physically active in the next 12 months. Again not at the mercy or the whims or the opinions of others, so neither agree nor disagree on that one. (participant 6, female)

This last quote illustrates another common problem identified, which is that for many questions, participants indicated strong verbal disagreement with a question, then indicated that they would endorse the ‘neither agree nor disagree’ response option on the questionnaire, as follows:

Q7. (motivation to comply). Generally speaking I want to do what my friends think I should. I tend to do what I want to do! So I’ll neither agree nor disagree! (participant 5, female)

Discussion

This first study identified a range of different types of problems encountered by participants when completing a 42-item TPB questionnaire about physical activity. The number of problems identified ranged from four to 18 problems identified per participant. Given the small number of participants, quantitative analyses were not conducted, but the examples quoted suggest that there were a range of different problems identified. The more straightforward problems identified arose because participants had problems comprehending the questions, due to question complexity. Problems of information retrieval were also present, when participants’ answers depended on other information that was not specified, or when participants did not feel they had appropriate knowledge or beliefs to answer the question adequately. The final category of problems arose because the participants answered different questions from the ones intended by the researchers.

Overall, the protocol analysis suggested that the participants in this study had a variety of problems with the TPB questionnaire they were asked to complete. However, by necessity, this first study employed a single TPB questionnaire, which was developed by a single research team, and which asked questions about a single behaviour of a single target population. For this reason, a second study was conducted, that employed a larger sample, and which asked questions about a single behaviour of a single target population (undergraduate students).

The particular aim of the second study was to assess whether similar problems were identified with this different TPB questionnaire. In addition, exploratory quantitative analyses were conducted,
to examine the extent of the problems identified, and to examine whether these problems were more common with some TPB measures than others. In particular, the frequencies of problems for attitudinal, normative or control variables were compared, as were the frequencies of problems for measures based on the belief-based measures that are derived from belief elicitation studies and which are often omitted from TPB studies, or for direct measures that are employed in all TPB studies.

Study Two

Method

Participants  The participants were 13 university students studying at a different university but in the same city as the target population of university students for which the TPB binge drinking questionnaire was developed. All participants were aged between 18–26 years; there were seven females and six males.

Materials  Participants completed a 62-item TPB questionnaire concerning binge drinking (questionnaire available from authors). This questionnaire assessed all 10 standard constructs of the TPB, including belief-based measures (see Table 4), derived from a belief elicitation study involving 83 participants (Norman & Conner, 2005), and was designed in accordance with the recommendations of Ajzen (1991) and Ajzen and Fishbein (1980). All items were phrased as incomplete statements or questions, with seven response options, numbered 1 to 7, with only the two extreme numbers labelled. In addition, there were two semantic differential questions that separately assessed the positive and negative features of the behaviour, designed to yield an indirect, ‘formula-based’ measure of ambivalence (see Conner & Sparks, 2002). A definition of binge drinking was given at the beginning of the questionnaire, based on drinking at least half the recommended maximum weekly number of units in a single session (see Murgraff, Parrott, & Bennett, 1999).

Procedure  The procedure employed was identical to that employed in the first study, with the exception that the instructions referred to ‘binge drinking’ rather than ‘increasing physical activity’.

Analysis  Chi-squared analyses were used to compare frequencies of problems for attitudinal, normative and control variables, and for belief-based versus direct measures.

Results

Segmenting and coding of think aloud transcripts  The first two authors independently segmented the 13 transcripts into material relating to each of the 62 TPB questions. Overall, there were 16 disagreements over segmenting, which were easily resolved following discussion. Given that there were $13 \times 62 = 806$ resulting segments, agreement in segmenting the text was over 98 percent (i.e. $100 \times 790/806$).

The same coding scheme was used as in the first study, and reasonable agreement was achieved between the first two authors, yielding a kappa based on the responses of all participants (corrected for chance agreement) of 0.60. As before, this level of agreement was arrived at after several iterations of independently coding transcripts, discussion of differences, then re-coding of the transcripts. One transcript was excluded, due to a large ($n=13$) number of items where there was no verbalization to code. The remaining 12 transcripts had only 10 items with no verbalizations made by participants between them. In these 12 transcripts, up to 14 problematic segments were identified, with no problems being identified for one participant. In contrast to the first study, there were relatively few problems with respondents re-reading questions and/or seriously floundering in answering them, although problems coded into the other categories were still fairly frequent (see Table 3). The frequency of problematic segments for each of these 11 constructs is shown in Table 4.

Distribution of problems identified by question type  Examination of Table 4 suggests that questions relating to the normative questions in the TPB, i.e. normative beliefs, motivation to comply and subjective norm, are particularly problematic. There are only seven questions assessing these constructs, but 20 of the 70 problems identified were in response to these few questions. Chi-squared analyses revealed that the frequency of problems associated with
Table 3. Frequency and type of problematic segments identified for the 12 respondents in Study Two

<table>
<thead>
<tr>
<th>Type of problem</th>
<th>Respondent number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No significant problems identified</td>
<td></td>
<td>56</td>
<td>52</td>
<td>55</td>
<td>60</td>
<td>56</td>
<td>48</td>
<td>57</td>
<td>56</td>
<td>58</td>
<td>57</td>
<td>57</td>
<td>62</td>
<td>674</td>
</tr>
<tr>
<td>Respondent re-read question, or seriously floundered in answering it</td>
<td></td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Respondent identified in self problems in understanding question</td>
<td></td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Questioned sensibleness of question</td>
<td></td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Respondent identified problem with how question was worded</td>
<td></td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Answered a different question from the one that was asked, or gave reasoning inconsistent with the answer given</td>
<td></td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Respondent had problems in comprehending/answering question, but were not themselves necessarily aware of this</td>
<td></td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Any of above problems</td>
<td></td>
<td>6</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>70</td>
</tr>
</tbody>
</table>
these questions was inconsistent with the null hypothesis of all items being equally problematic ($\chi^2 = 23.1$, d.f. = $1$, $N = 744$, $p < .0001$).

These analyses were repeated for each category of problem identified, showing that problems relating to how sensible participants found the questions were not more frequent than would be expected by chance ($\chi^2 = 0.13$, d.f. = $1$, $N = 744$, $p = .72$). However, problems relating to floundering ($\chi^2 = 5.55$, d.f. = $1$, $N = 744$, $p < .019$) and, in particular, answering a different question from the one asked ($\chi^2 = 28.7$, d.f. = $1$, $N = 744$, $p < .0001$) were much more frequent than the null hypothesis of all items being equally problematic would predict.

Table 4 also shows that there are 40 questions assessing belief-based constructs which yielded 43 problematic responses, in contrast to 27 problems yielded by the remaining 22 questions. Chi-squared analyses showed that the frequency of problems associated with belief-based and direct questions was consistent with the null hypothesis of all items being equally problematic ($\chi^2 = 0.32$, d.f. = $1$, $N = 744$, $p < .57$). These analyses were repeated for each category of problem identified, showing that problems relating to floundering were more frequent than would be expected by chance for direct measures compared with belief-based measures ($\chi^2 = 7.80$, d.f. = $1$, $N = 744$, $p < .01$). However, problems relating to how sensible participants thought the questions were ($\chi^2 = 0.06$, d.f. = $1$, $N = 744$, $p = .63$) and answering a different question from the one asked ($\chi^2 = 0.06$, d.f. = $1$, $N = 744$, $p = .81$) were consistent with the null hypothesis of all items being equally problematic.

Examples of the problems identified

**Participant re-read question, or seriously floundered in answering it**

As was the case with the physical activity questionnaire, floundering apparently arose because the participants thought their answers depended on further information that was not specified in the question:

Q46. (perceived behavioural control). If I wanted to I could easily engage in a binge drinking session over the next week. Er might do. It depends on whether I go out or not. Er I will put three. If I wanted to I could easily engage in a binge drinking session over the next week. Yeah. seven. (participant 2, female)

In contrast to the physical activity questionnaire however, most problems involving re-reading or floundering with this questionnaire were apparently due to questions that involved the phrase ‘should or should not’ in the stem:

Q49. (subjective norm). People who are important to me think I should or should not engage in a binge drinking session over the next week. Think I should not or think I should? People who are important to me … People who think binge drinking is not that great
but also want me to have so fun so probably towards think I should not, put 3. (participant 3, female)

Q24. (motivation to comply). With regard to binge drinking, how much do you want to … How much do you want to do with your friends that think you should? Not at all … Have I read that right? With regard to binge drinking, how much do you want to do with your friends … What your friends think you should? I generally do what I want to do and I don’t always do what people think I should do. So probably 3, not at all. (participant 3, female)

Participant questioned how sensible the questions were Again, in common with the physical activity questionnaire, participants often questioned how sensible the questions they were answering were when they thought their answers depended on further information that was not specified in the question:

Q11. (behavioural belief). Engaging in a binge drinking session over the next week would cost a lot of money erm … It depends on what I drink erm … If I am drinking Martinis, which is what I like. Yes. If I’m drinking horrible alcopops that stain my tongue. Then no. [laughs] Erm … It does cost money but to me not a lot of money. It costs money but not loads and loads that I’m not prepared to spend because I quite enjoy it, so it would cost money. Say 3. Because even though it costs money it doesn’t cost a lot of money. (participant 6, female)

Q17. (behavioural belief). Engaging in a binge drinking session over the next week will increase my confidence. Maybe for the night, but not for the next day. [pause] … (participant 8, female)

Participants were sometimes not clear exactly to what the questions were referring, in line with responses to the physical activity questionnaire. In the present study, a number of participants were unsure whether the questions referred to the period during which they might be binge drinking, or over a longer period of time:

Q13. (behavioural belief). Engaging in a binge drinking session over the next week will increase my confidence. Maybe for the night, but not for the next day. [pause] … (participant 11, female)

Q56. (attitude). Engaging in a binge drinking session over the next week would be … It would be pleasant at the time but it’s the next day that’s the problem. So I’d say the binge drinking session, which is talking about the actual session and not the next day, is pleasant because … So that’s probably a 2 towards pleasant. (participant 3, male)

A similar problem occurred when participants were asked about what important other people would think about them binge drinking, due to participants believing that different people would have different opinions:

Q23. (normative belief). My friends think I should/should not engage in a binge drinking session over the next week. Depends which friends you’re talking about. University friends, definitely are up for binge drinking. But my more personal, closer friends probably aren’t. So they … I’d probably say that those people I socialize most with, and they think I should not so give that a 3 towards binge should not. (participant 3, female)

Q42. (subjective norm). People who are important to me would approve or disapprove of me engaging in a binge drinking session over the next week. Erm does that mean friends or parents? Parents always disapprove but your friends are like … they do the same so … Erm I will put 3. (participant 2, female)

Participants also questioned how sensible the questions they were answering were when they felt they had already answered the same question:

Q19. (behavioural belief). Engaging in a binge drinking session over the next week would make me feel ill. That’s the same question as a hangover really. (participant 10, male)

Q51. (intention). I will engage in a binge drinking session over the next week. Yeah that one is the same as do you intend engaging in a binge drinking session over the next week. So seven. (participant 6, female)

Participant answered a different question from the one that was asked, or gave reasoning inconsistent with the answer given As with the physical activity questionnaire, several participants answered a different question from the one that was asked, using information that the question did not seek. In the following two examples, the participants responded on the basis of the social situation anticipated, rather than binge drinking per se, and misconstrued ‘control’ as control over behaviour having been drinking, rather than control over amount of drinking:

Q5. (behavioural belief). Engaging in binge drinking sessions over the next week would be fun. Engaging in? As in my friends tend to do it and I would be in

FRENCH ET AL.: THINKING ALOUD AND TPB QUESTIONNAIRES

© 2007 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.
that same social party of people so I suppose I would be having fun, but I just don’t drink. So, I don’t do the binge drinking, so it’s quite likely that I would be involved. (participant 9, female)

Q44. (perceived behavioural control). How much control do you have over whether or not you engage in a binge drinking session over the next week. Well I don’t get naked and run around. (participant 11, male)

As in the previous study, several participants answered questions about what other people wanted them to do by referring to whether this mattered to them:

Q42. (subjective norm). People who are important to me would approve/disapprove if I engaged in a binge drinking session next week. I don’t really care so I’ll go for the middle. (participant 1, male)

The last quotation illustrates the problem of participants indicating verbal agreement or disagreement with a question, then indicating that they would endorse the ‘neither agree nor disagree’ response option on the questionnaire:

Q35. (perceived inhibiting/facilitating power). Experiencing sociable pressures to drink would make engaging in binge drinking session over the next week. Erm … My friends don’t put a lot of pressure on me to drink, they put pressure on me to go out because we have a laugh when we go out erm … And if friends are putting social pressure on me to drink I probably wouldn’t listen to them because I’ve got … I make my own choices in what I drink erm … And I actually think people that are giving into social pressures to drinking is bad. So it would not really affect it whatsoever. So four. (participant 6, female)

Discussion

In line with the first study, this second study identified a range of different types of problems encountered by participants when completing a 62-item TPB questionnaire about binge drinking. Although no problems were identified for one participant, a mean of 5.8 problems was identified for the 12 participants as a whole. In contrast to the first study, relatively few problems with participants floundering or re-reading questions were noted, although problems with participants questioning how sensible the questions were, or drawing upon irrelevant material in answering them, were still fairly common.

Quantitative analyses revealed that questions relating to perceived normative influence were particularly problematic. Relative to other questions, participants were more likely to answer a different question from the one asked for normative questions. By contrast, participants completing direct measures were more likely to re-read these questions or flounder when answering them, relative to belief-based measures.

The examples quoted suggest that there was a range of different problems identified. These included problems of information retrieval, when participants’ answers depended on other information that was not specified, or participants answering different questions from those intended, or selecting response options that were not congruent with the reasoning verbalized for that question.

Overall, the findings of the second study are in broad agreement with those of the first study. The second study employed a questionnaire which was developed by a different research team, and which asked questions about a different behaviour of a different target population. The similarity of the problems encountered while completing both TPB questionnaires suggests that these problems are not due to the quirks of a specific questionnaire, but are more general.

Overall discussion

The first point that should be made when considering the two studies together is that each participant did not have problems with most of the questions that they answered. In the first study, a mean of 9.7 problems was identified per participant in completing a 42-item questionnaire about physical activity; in the second study, a mean of 5.8 problems was identified per participant in completing a 62-item questionnaire about binge drinking. It is entirely possible, however, that some problems were encountered that it was not possible to detect reliably using protocol analysis. This technique relies on participants verbalizing their thoughts: to the extent that participants do not do so, problems that are not verbalized cannot be detected.

When discussing the problems identified, it may be worth keeping in mind two distinctions. First, some problems may well be encountered when people are asked to complete any kind of questionnaire, whereas some problems may arise uniquely or more frequently with TPB questionnaires, due to the nature of the questions asked. Second, some problems may be resolved or ameliorated by
piloting: once identified, the problems can be tackled. By contrast, some problems may be less tractable. In the following paragraphs, we discuss the types of problems that were quoted as examples, in the light of the two distinctions just outlined.

**Participant re-read question, or seriously floundered in answering it**

The first type of problem identified involved participants re-reading a question or seriously floundering in answering it, due to the complexity of the question. In Study One, questions that required participants to disagree with a negatively phrased statement or respond to a hypothetical question caused problems; in Study Two, a question that included the phrase ‘should or should not’ caused problems. The first problem should be readily soluble by positively phrasing the question. The problem identified in Study Two should be solved by omitting ‘or should not’, but keeping the response anchors ‘should’ and ‘should not’. The second problem in Study One, involving responding to a hypothetical question, may be less easily resolved. The question: ‘if I had a lack of interest in being more physically active in the next 12 months it would make it difficult for me to be more physically active’ was derived from responses in a belief elicitation study about what would make it difficult to be more physically active. Where belief elicitation studies yield a number of responses about lack of motivation as a barrier to performing a behaviour, questionnaires produced in accordance with Ajzen’s (2002) current recommendations will consequently contain problematic items of this type.

It is noteworthy that problems involving participants re-reading a question or floundering in answering it occurred more frequently in Study One than in Study Two. This may be a statistical fluke, due to the small numbers of participants involved. However, assuming that there is a real difference in problems of this nature between the two studies, two obvious explanations arise, which have different implications.

Study One used a format which differed from that recommended for TPB questionnaires, as it contained statements which participants were required to agree or disagree with on a five-point scale, instead of the recommended seven-point scales, with each question having different anchors. Although this response format differs from recommendations, similar formats are common in TPB studies (e.g. Abraham, Henderson, & Der, 2004; Drossaert, Boer, & Seydel, 2005). If the format was responsible for the problems identified in Study One, then the obvious solution is to use the recommended format.

However, a more likely explanation in our opinion concerns the nature of the samples in the two studies: participants in Study One were recruited from the general population, whereas the participants in Study Two were university undergraduates. It seems reasonable to suppose that more educated participants would have fewer problems with comprehending questions, especially where questions involve hypothetical states of affairs, as is the case with all TPB questionnaires. Evidence for this viewpoint comes from the meta-analytic finding that associations between constructs were stronger for student samples than other samples (Armitage & Conner, 1996).

Another kind of problem that involved participants re-reading a question or seriously floundering in answering it appeared to be due to their answers depending on further information, which was not specified in the question. In the first study, questions concerning the impact of increasing physical activity on significant other people caused problems, due to the questions not specifying the activity and whether these significant others would also be involved. However, the second study yielded fewer problems of this nature, despite ‘binge drinking’ being considered, rather than drinking particular sorts of drinks or binge drinking in a particular context. This was important in a few situations, for example where control over whether or not the participant binge drinks depends on whether they go out or not. It is difficult to see how any questionnaire could specify the full range of alternative scenarios when asking these types of questions without becoming extremely lengthy.

**Participant questioned how sensible the questions were**

A similar kind of issue also arose when participants queried how sensible the questions they were answering were, as they felt their answers depended on further information that was not specified in the question. In the first study, answers to a question about spending time with family depended on whether it referred to the general case, or the specific case of spending time with family while being physically active. In the second study, answers to questions about the consequences of binge drinking depended on what drinks were being consumed and...
who the participant thought they might be drinking with. As before, it would not be feasible for any questionnaire to specify the full range of alternative scenarios when asking these types of questions. However, these problems do highlight that even clearly defined behaviours such as ‘drinking more than five pints (or 10 shorts/glasses of wine) in a single session for men’ can be open to a variety of interpretations: the behaviours that are defined according to public health criteria do not exactly map on to single scenarios for individual participants.

Participants also queried whether the questions they were answering were sensible when they were unclear how to interpret these questions. In the first study, having ‘spare time’ could be construed as spending time being physically active and/or time spent with family. In the second study, participants were unclear whether questions asking about the consequences of binge drinking referred to the drinking session itself, or the days following the session. The issue of whether questions about the perceived consequences of binge drinking should focus on the session itself or the longer-term consequences is one that the current procedure for eliciting salient beliefs in the TPB does not address. Given the theoretical and empirical work on the implications of current and future time perspectives for health-related behaviours (e.g. Orbell, Perugini, & Rakow, 2004; Zimbardo, Keough, & Boyd, 1997), future TPB elicitation studies might consider this distinction further.

In the second study, participants also found it difficult to answer questions about the views of categories of people, such as friends or ‘people who are important to me’, as they thought different people within these categories would have different views. In belief elicitation studies, participants are asked to cite people who would approve or disapprove of the particular behaviour being studied, and often give categories of people. Consequently, questions asking about the views of groups of people whose individual members may not agree with each other, as suggested by the TPB procedure for developing questionnaires, are likely to continue to prove problematic across TPB studies.

The third category of problem that led participants to query whether the questions they were answering were sensible arose when they felt there was only one sensible or coherent answer to be given. In the first study, this was brought about by participants noting that the answers to questions were implied by previous questions. In the second study, participants felt that some questions were so similar to previous questions that they meant the same thing. This problem is likely to occur in all questionnaires that use scales comprising multiple items assessing related ideas, rather than being unique to the TPB. It is worth noting, however, that the original guidelines for developing TPB measures (Ajzen & Fishbein, 1980) recommended a single item measure of intention, whereas it was subsequently recognized that this is less than ideal (see Sutton, 1998). More recently, intention scales formed from multiple items have become both the norm and the current recommendation for TPB measures (Ajzen, 2002).

**Participant answered a different question from the one that was asked, or gave reasoning inconsistent with the answer given**

The third category of problems identified concerned participants answering a different question from the one that was asked, or giving reasoning inconsistent with the answers they gave. Participants answering a different question from the one asked was apparent when they appeared to use information that was not sought by the question in formulating their answers. In both studies, it was apparent that many participants answered normative belief items, concerning whether specified people would approve of the particular behaviour, with reference to their motivation to comply with these people, usually stressing that the opinions of these other people were not important. This is clearly a problem that is of particular relevance to the TPB, given that normative beliefs and motivation to comply are hypothesized to be distinct constructs that combine in a multiplicative way to form overall subjective norms.

The final type of problem identified concerned participants giving answers inconsistent to their verbalized reasoning. A particularly common instance of this occurred when participants (often strenuously) disagreed that they were influenced by the opinions of their important others, yet chose the middle response option, not the extreme ‘disagree’ response option. This may be a manifestation of the common ‘response acquiescence’ phenomenon (see Meier, 1994), and so general to most types of questionnaire. On the other hand, this problem was particularly common in responses to motivation to comply questions, where participants appeared to translate their expressed indifference to the opinions of others into choosing the middle response.
option, rather than the extreme option that would appropriately reflect their stated views.

**Quantitative analyses**

Having discussed the types of problems that were quoted as examples, the quantitative analyses reported in Study Two allow the extent of these various problems to be established. The most striking finding about the quantitative analyses was the much higher prevalence of problems with questions about normative influence (i.e. normative beliefs, motivation to comply and subjective norm) than with questions relating to the other TPB constructs. Although quantitative analyses were only conducted for the data gathered in the second study, it is apparent in Table 2 that these items were also among the most problematic in the first study. It is a commonplace observation in the TPB literature that the subjective norm construct has much less predictive power than do the attitude and perceived behavioural control constructs (see Armitage & Conner, 2001; Godin & Kok, 1996). Closer scrutiny of the problems encountered here may shed some light on why this is the case.

Three types of problem that appear specific to TPB normative questions have already been discussed: (a) participants indicated that groups such as ‘friends’ or ‘people who are important to me’ are comprised of sub-groups who have different opinions, making it difficult to give a single answer; (b) participants answering normative belief questions by reference to their motivations to comply; (c) answering all types of normative questions, but particularly those concerning motivation to comply, by verbally indicating disagreement, then choosing the middle response option. The first of these problems might be addressed by employing procedures akin to those employed to assess ambivalence (e.g. asking about whether those friends who approve of binge drinking approve strongly, and those who disapprove are strong in their disapproval (see Conner & Sparks, 2002)). The second problem appears to indicate that being consciously influenced by other people is not seen as socially acceptable or desirable. The extent to which this is the case is an empirical question, but does call into question the usefulness of normative belief questions if they are not acceptable to a sizeable minority of potential respondents. An alternative approach might be to have the motivation to comply questions first, which might lead respondents to consider they have already indicated their views on this topic, and so do not need to do so again (Schwarz, Strack, & Mai, 1991). The third problem may again indicate that certain responses are not socially acceptable or desirable, or may simply reflect problems in formatting responses.

In total, the questions assessing normative aspects of the TPB appear to be particularly problematic for a number of reasons, and these constructs are not as predictive of intentions as are other TPB constructs (Armitage & Conner, 2001).

Several lines of research are currently being conducted to widen out the conception of normative influence within the TPB, by including constructs such as descriptive norms, relating to perceptions of what important others do, not what they think one should do (see Donald & Cooper, 2001; Rivas & Sheeran, 2003; Terry, Hogg, & White, 1999). In addition to these developments, the present findings suggest that normative constructs may be more predictive if greater attention is paid to resolving the problems that people have when completing them.

**Concluding comments**

We have reported the first data yielded to our knowledge by cognitive interviewing with participants completing TPB questionnaires, concerning physical activity or binge drinking. The findings reported here should be of general interest to researchers who utilize TPB questionnaires, not least because the studies reported used questionnaires that were developed on the basis of belief elicitation studies, as recommended (Ajzen, 2002; Ajzen & Fishbein, 1980). The two questionnaires also addressed different behaviours with different populations, and were developed by different research groups. There are, nevertheless, several limitations of this study, due to small sample size, consequent limitations in statistical power and lack of a priori hypotheses. However, we see the findings reported here as being exploratory, suggesting areas where future, better-powered studies might test hypotheses that the current research has generated.

At the moment, we consider our pattern of findings as potentially problematic areas that warrant further attention, rather than definitive. More intensive qualitative analysis may be a useful way forward, with the aim of understanding more fully the phenomenology of questionnaire completion. In addition, more systematic research with larger sample sizes will need to be conducted before robust conclusions about the validity of TPB questionnaires can be
reached. Furthermore, given the nature of the samples used, our data do not have anything useful to say about whether the problems we have identified are more common for populations such as those with little education or ethnic minorities. Future research should explicitly attempt to include disadvantaged groups; it is entirely possible that the problems we have identified here may be more common when such populations are studied.

We believe we have not only identified problems that may affect all TPB questionnaires, particularly related to questions assessing normative influence, but also given a demonstration of how protocol analysis can yield useful information about questionnaires developed according to detailed and much-used guidelines. Given that most TPB questionnaires are used only once or a few times with a specified population and behaviour, a thorough psychometric evaluation is usually not feasible. However, with the increasing use of intervention studies that are based on changing TPB constructs (e.g. Williams et al., 2004), large sums of money may be spent on trials that rely on TPB questionnaires to assess process, but which have no evidence of validity. The use of the procedures presented here may be a useful but inexpensive way to gather some data on acceptability of TPB questionnaires before being employed in intervention studies.

In conclusion, we have identified some problems that appear to be common to two TPB questionnaires that were independently developed, in line with current guidelines (Ajzen, 1991, 2002; Ajzen & Fishbein, 1980). Given the widespread use of the TPB in health psychology, systematic problems with questionnaires developed according to these guidelines should be avoided. Future studies should determine the extent of these problems, and might consider how they can be resolved.

References


**Author biographies**

**DAVID FRENCH,** is a lecturer in health and exercise psychology at the University of Birmingham. His research interests include health behaviour change, responses to health risk information, and issues in psychological measurement.

**RICHARD COOKE,** is a lecturer in health psychology in the Department of Psychology, Aslon University. His research interests include using social cognition models to predict binge-drinking behaviour, issues in psychological measurement and understanding how people process health information.

**NIKKY McLEAN,** was a Research Associate at the University of Cambridge while this research was conducted. She is now head Practitioner Dietition at East Cambridgeohine and Fenland PCT.

**MERRONEY WILLIAMS** is an under-graduate psychology student at Sheffield Hallam University.

**STEPHEN SUTTON** is professor of Behavioural Science at the University of Cambridge. His research programme focuses on risk communication and health behaviour change. He has a particular interest in theoretical and methodological issues.