Understanding Society Innovation Panel Wave 3:
Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd, Jonathan Burton, Emily Gilbert, Annette Jäckle, Stephanie McFall, SC Noah Uhrig

Institute for Social and Economic Research
University of Essex

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Understanding Society Innovation Panel Wave 3:
Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd, Jonathan Burton, Emily Gilbert, Annette Jäckle,
Stephanie McFall, SC Noah Uhrig

Abstract

This paper presents some preliminary findings from the Wave 3 Innovation Panel (IP3) of *Understanding Society*: The UK Household Longitudinal Study. *Understanding Society* is a major new panel survey for the UK. In April 2010, the third wave of the Innovation Panel was fielded. This paper describes the design of IP3, the experiments carried and the preliminary findings from early analysis of the data. The main design features of *Understanding Society* are outlined and the design and conduct of IP3 described. The results of methodological experiments carried at IP3 are reported and the impact of IP3 on the design of the main survey is reviewed.

Understanding Society Innovation Panel Wave 3: Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd, Jonathan Burton, Emily Gilbert, Annette Jäckle, Stephanie McFall, SC Noah Uhrig

Non-technical summary

The *Understanding Society* survey includes what is known as an 'Innovation Panel' sample. This sample of 1500 households is used to test different methods for conducting longitudinal surveys in order to produce the highest quality data. The results from the Innovation Panel provide evidence about the best way to conduct a longitudinal survey which is of relevance for all survey practitioners as well as influencing decisions made about how to conduct *Understanding Society*. This paper reports the results of the methodological tests carried out at wave 3 of the Innovation Panel in the spring of 2010 and comments on how the results from the experiments influenced decisions made for the main survey.

The methodological tests included an experiment testing the effects of different incentives offered to respondents in advance of fieldwork on response rates. Further tests examined different ways of designing survey questions, including different formats and question wording for asking attitude questions, for asking about changes in people's circumstances, and for collecting information about household wealth. The survey in addition tested different methods of measuring cognitive functioning.

1   Introduction

This paper presents early findings from the third wave of the Innovation Panel (IP3) of *Understanding Society*: The UK Household Longitudinal Study (UKHLS). The UKHLS is a major new panel survey for the UK. The first wave of data collection, covering two years, has been completed (January 2009 to March 2011) and the first year of data is available from the UK Data Archive (http://www.esds.ac.uk/findingData/usoc.asp). The full Wave 1 data-set, which includes the ethnic minority boost is available since November 2011. Data for the first two waves of the Innovation Panel are available from the Data Archive.

One of the features of the UKHLS, alongside the large sample size (40,000 households at Wave 1), the ethnic minority boost sample and the collection of bio-markers, is the desire to be innovative. This has been a key element of the design of the UKHLS since it was first proposed. Part of this drive for innovation is embodied within the Innovation Panel (IP). This panel of some 1500 households were first interviewed in the early months of 2008. The design in terms of the questionnaire content and sample following rules are modelled on the UKHLS. The IP is used for methodological testing and experimentation that would not be feasible on the main sample. The IP is used to test different fieldwork designs, new questions and new ways of asking existing questions.

The second wave of the Innovation Panel (IP2) was carried out in April-June 2009. Working Papers which cover the experimentation carried out in both IP1 and IP2 are available from the *Understanding Society* website.[1] The IP1 and IP2 data are held at the UK Data Archive and are made available via the Economic and Social Data Service.[2] The data from IP3 and IP4 are due to be deposited and available to researchers in the summer of 2012. In March 2010, IP3 was fielded. This paper describes the design of IP3, the experiments carried and some preliminary findings from early analysis of the data. Section 2 outlines the main design features of the UKHLS. Section 3 describes the design and conduct of IP3. Section 4 then reports on the experiments carried at IP3. Section 5 reviews the impact on IP3 on the design of the main survey and Section 6 concludes the paper by outlining plans for IP4.

2   *Understanding Society*: the UKHLS

The UKHLS is an initiative of the Economic and Social Research Council (ESRC) and is one of the major investments in social science in the UK. The study is managed by the Scientific Leadership Team (SLT), based at ISER at the University of Essex and including members from the University of Warwick and the Institute of Education. The fieldwork and delivery of the survey data is undertaken by the National Centre for Social Research (NatCen). The UKHLS aims to be the largest survey of its kind in the world. At the end of Wave 1 the sample size was 39,805 households, covering 101,087 individuals. The sample covers the whole of the UK, including Northern Ireland and the Highlands and Islands of Scotland. The UKHLS provides high quality, longitudinal survey data for academic and policy research across different disciplines. The use of geo-coded linked data enables greater research on

---

[1] http://research.understandingsociety.org.uk/publications/working-paper/2008-03.pdf
http://research.understandingsociety.org.uk/publications/working-paper/2010-04.pdf
[2] http://www.esds.ac.uk/findingData/snDescription.asp?sn=6849

neighbourhood and area effects, whilst the introduction of bio-markers and physical measurements (Waves 2 and 3) opens up the survey to health analysts.

The design of the UKHLS is similar to that of the British Household Panel Survey (BHPS) and other national panels around the world. In the first wave of data collection, a sample of addresses was issued. Up to three dwelling units at each address were randomly selected, and then up to three households within each dwelling unit were randomly selected. Sample households were then contacted by NatCen interviewers and the membership of the household enumerated. Those aged 16 or over were eligible for a full adult interview, whilst those aged 10-15 were eligible for a youth self-completion. The adult interviews were conducted using computer-assisted personal interviewing (CAPI) using lap-tops running the questionnaire in Blaise software. Adults who participated in the UKHLS were also asked to complete a self-completion questionnaire, in which questions thought to be more sensitive were placed. The adult self-completions at Waves 1 and 2, and the youth self-completions, were paper questionnaires. From Wave 3 onwards the adult self-completion instrument was integrated into the Blaise instrument and the respondent used the interviewer's lap-top to complete that portion of the questionnaire themselves (Computer-Assisted Self-Interviewing, CASI).

In between each wave of data collection, sample members are sent a short report of early findings from the survey, and a confirmation-of-address slip, to allow them to confirm their address and contact details. Before each sample month is issued to field for a new wave, each adult is sent a letter which informs them about the new wave of a survey, includes a token of appreciation in the form of a gift voucher and also includes a change-of-address card. Interviewers then attempt to contact households and enumerate them, getting information of any new entrants into the household and the location of anyone who has moved from the household. New entrants are eligible for inclusion in the household. Those who move, within the UK, are traced and interviewed at their new address. Those people living with the sample member are also temporarily eligible for interview. More information about the sampling design of the UKHLS are available in Lynn (2009).[3] From Wave 2, the BHPS sample has been incorporated into the UKHLS sample. The BHPS sample are interviewed in the first year of each wave.

3   Innovation Panel Wave 3: Design

Unlike IP2, which used a mixed-mode design, IP3 was completely face-to-face, using CAPI interviewers. The fieldwork ran from the 12th April to the 27th June. The initial issue period was extended because of slower-than-expected progress. Prior to the survey going into the field there were seven one-day briefings for the interviewers. The briefings were conducted by NatCen researchers, with staff from ISER contributing to provide information about the study and to talk in more detail about the experiments. The briefings were held in Bristol, Derby, Glasgow, Leeds, Liverpool and London (2 briefings). In total, 120 interviewers were briefed to work on IP3. A dozen of these interviewers also attended a de-brief session in London on 30th June. The questionnaires used at IP3 are available from the *Understanding Society* website.[4]

---

[3] http://research.understandingsociety.org.uk/publications/working-paper/2009-01.pdf
[4] http://data.understandingsociety.org.uk/questionnaires/wave-3

a.   Call for experiments

IP3 was the first time the Innovation Panel was open for researchers outside the scientific team of the UKHLS to propose experiments. A public call for proposals was made.. Eleven proposals were received, four were initially accepted although one was eventually dropped and another was held over for IP4. Submissions came from within ISER (three), ISER in collaboration with other researchers (two) and from outside ISER completely (six). Of those that were external to ISER, two were from the USA and the other four were from UK-based researchers. The six external proposals were from six different organisations, including universities, government department and private companies. The eleven proposals were reviewed by a panel which included two ISER-based members of the UKHLS scientific leadership team, one senior researcher from the fieldwork agency and one external expert from the University of Southampton. In future waves of the IP, the call for experiments will continue to be open to researchers outside the UKHLS scientific team.

b.   Sample

The issued sample at IP3 comprised those household who had been productive at IP2, plus households which had been productive at IP1, but not IP2. IP2 was a mixed-mode survey and so some of the household had been contacted by telephone. At IP3 we wanted to see whether we could recover any of the households that we had lost at IP2, through not being able to contact them or 'soft' refusals. Altogether 1,513 households were issued at IP3; 1,119 were previous-wave productive households, 141 were non-contacts at IP2 and 253 were refusals at IP2.

c.   Interview schedule

The questionnaire at IP3 followed the standard format used in the two previous Innovation Panels as well as the main-stage of the UKHLS. The interview included:
- Household roster and household questionnaire: 15 minutes per household
- Individual questionnaire: on average 37.5 minutes for each person aged 16 or over
- Adult self-completion: around 7-10 minutes, paper questionnaire
- Youth self-completion: 10 minutes for each child aged 10-15 years
- Proxy questionnaire: 10 minutes for adults ages 16 or over who are not able to be interviewed.

The individual adult interview was longer than in previous waves because of the addition of a cognitive functioning module. This included three tasks; a measure of verbal fluency (FAS test), prospective memory and a measure of working memory (Serial-7s). Section 4a contains more information about the cognitive function modules.

Some parts of the IP3 interview were recorded, using the lap-tops. This was done with permission of the respondent. The recordings were primarily around the experimental

content of the questionnaire, and were taken to enable researchers to investigate the processes by which respondents came up with their answer. Around 72% of those who participated agreed to the sound recording.

d.   Response rates

This section sets out the response rates for IP3 as a whole. Section 4e describes the effect of incentives on response rates. There were 1,513 households issued to field, 50 of those were found to be ineligible at IP3, whilst there were 75 'split-offs', where one or more individuals had left the issued household and moved elsewhere, leaving sample members in the original household. This gives an eligible sample of 1,538 households. Table 1 shows the household response for the IP3 eligible sample as a whole, and split by those who were productive at IP2 and those households that were not.

Table 1: Household response at IP3

	IP3 eligible sample %	IP2 productive %	IP2 unproductive %
Productive (household and at least one individual)	66.1	77.6	32.7
Non-contact	4.3	2.7	8.9
Untraced movers	6.5	2.9	17.0
Refusals	17.5	12.7	31.5
Other unproductive	5.6	4.1	9.9
n	1,538	1,144	394

Overall, around two-thirds of households were productive at IP3, that is, there was a household interview and at least one individual adult interview. There were a small number of households where a household-level interview was done, but no individual interviews. These are classified as "other unproductive". Amongst those households that were productive at IP1 and IP2, the response was much higher, with over three-quarters being productive. For those households that were productive at IP1 but not at IP2, nearly one-third were productive at IP3. This demonstrates the value of issuing previous-wave unproductive cases on a longitudinal survey. The relatively high proportion of the issued sample that ended up as untraced movers (6.5%) underlines the importance of getting as much contact information as possible. The untraced mover rate was much lower among those who were interviewed at IP2 (2.9%) than those who were unproductive at IP2 (17.0%). Those households that were non-contacted at IP2 were much more likely to be untraced movers at IP3. Table 2 below shows the individual-level response rate within productive households.

Table 2: Individual response within productive households at IP3

	Individuals in productive households %
Full interview	81.7
Proxy interview	6.8
Non-contact	2.2
Refusal	6.6
Other non-response	2.5
n	1,972

4 Experimentation in IP3

The experiments carried on IP3 covered both fieldwork procedures and measurement in the questionnaire. There were some new experiments and some which were the longitudinal continuation of experiments carried at IP1 and/or IP2. This section outlines the experiments carried at IP3, briefly explaining the reasons for carrying them, describing the design of the experiment and giving an indication as to the initial results from early analysis of the data. In addition, a description of the cognitive function measures is included. This was not implemented experimentally for IP3, but was carried to see whether it was feasible to collect these measures as part of a broader social survey.

a. Cognitive function measures

Wave 3 of the Innovation Panel contained three cognitive measures: prospective memory; a verbal (phonological) fluency measure (FAS test); and Serial 7 subtraction, which measured working memory. The purpose was to gain experience with this domain using brief measures and to explore selected methodological issues. This overview describes the measures and reports on item non-response and patterns of response in relation to age.

i. Prospective memory

This is the ability to remember to do something. The interviewer explained that at some point they will hand the respondent a piece of paper and a pencil and the respondent should at that point write his or her date of birth on the top left-hand corner of the paper. If the respondent does not do the task within 5 seconds of being handed the paper, the interviewer prompts the respondent with "You were going to do something when I gave you the paper and pencil. Can you remember what it was?". The response codes were 1) Wrote date of birth in top left-hand corner, 2) Wrote date of birth somewhere else, 3) Wrote something else in top left-hand corner , or 4) Did something else.

ii. Phonemic fluency

This task measures the ability to generate, access and produce words starting with a certain letter in one minute. Respondents are randomly assigned to words beginning with F, A, or S. The interviewer begins the timing when the respondent produces the first word. If the respondent says, "I can't think of any more," the interviewer gives encouragement—"Keep trying." The interviewer records the number of correct and incorrect words. The interviewer also records whether others were present and who was present.

iii.   Serial 7 Test

Working memory was assessed using the Serial 7 subtraction test. Working memory refers to short-term processes used to store and make use of information in more complex tasks. In this test the interviewer asked the respondent to subtract 7 from 100, and continue subtracting 7 from each subsequent answer for a total of 5 trials. In this question series, the respondent must remember the answer from the previous subtraction. The interviewer recorded the number, or whether the respondent said that they didn't know, or refused. For the series, the interviewer also recorded whether the respondent used paper and pencil or other aids. The interviewer recorded the presence of any other person and who that was.

Scores could range from 0 (all wrong) to 5 (all correct). Each subtraction was scored independently. That is, if a respondent made a mistake on the first subtraction, but gave correct answers for subsequent subtractions, the score would be 4. Respondents who refused to perform the test at the outset or who began the test and refused mid-way through were assigned missing values.

iv.   Item non-response

Two respondents did not get to the cognitive module. The table below is based on the 1,619 respondents who did get to the cognitive module. The amount of item non-response was small, less than one percent for the prospective memory and FAS tests, and around 3% for the Serial 7 test.

Table 3: Item non-response for cognitive measures

	Prospective memory		FAS		Serial 7	
	n	%	n	%	n	%
Refused	12	.7	7	.4	24	1.48
Don't Know	4	.25	7	.4	34*	2.10

1619 individuals responded to the cognitive function module
* Frequency with don't know on the first subtraction item

Item non-response is often related to non-response on other cognitive measures or to how they perform on those tests if they agree to participate. To illustrate, we examine the 24 respondents who refused one or more of the subtraction tasks in the Serial 7 and their performance on the FAS. One-fifth of those who refused the Serial 7 task also refused the FAS test, whilst just under one-third scored zero on the FAS test. In addition, 7 did not correctly perform the prospective memory task and 7 refused both tasks.

This suggests that although the amount of item non-response on the cognitive measures is small, it may be biased towards respondents who perform less well or who fear that they may not do well. The different cognitive function measures are positively correlated, with stronger associations between the Serial 7 and the FAS test (0.2792) than between Serial 7 and prospective memory (0.1847) or the prospective memory and the FAS test (0.1589). Future research should assess the construction of a composite cognitive function measure.

v.   Association with age

Beginning with age 25, age was categorized in decades. The exceptions were the youngest category (16-24 years) and the oldest category (85 or older). Table 4, below, presents mean scores for the FAS and Serial 7 and the percentage performing the prospective memory task correctly by age category.

Performance of the prospective memory task is relatively stable until age 64, shows some decline for age 65-74 and then substantial decline. FAS scores peak between 35 and 64. There is a gradual decline after age 64 and more after age 85. The Serial 7 scores are largely stable until age 84.

Table 4: Cognitive measures by age category

	Prospective memory		FAS		Serial 7	
	%	n	Mean	SD	Mean	SD
16-24	78.7	118	10.3	4.7	4.1	1.2
25-34	79.4	150	11.7	5.3	4.3	1.3
35-44	84.1	238	12.5	6.1	4.2	1.3
45-54	81.0	251	13.2	5.8	4.2	1.2
55-64	77.5	200	13.2	5.4	4.3	1.3
65-74	65.4	153	11.5	5.8	4.0	1.4
75-84	49.6	71	10.5	5.8	4.0	1.5
≥ 85	47.4	18	8.2	5.8	3.6	1.7
total N		1605	1597		1581	

vi.   Methodological opportunities

Participants were asked if their interview could be audio-taped, and 82% of respondents agreed to recording of the full interview or the cognitive portion. While this type of data will not be released under the End-User License, it would permit examination of more detailed cognitive indicators related to the FAS, as well as examination of a range of interviewer and respondent behaviours.

Approximately a quarter of respondents were not alone when they completed the cognitive module. Researchers could examine the effects of assistance from others as well as whether there are order effects in which persons interviewed later have higher cognitive scores.

b.   Measures of change

As part of the ongoing programme looking at 'measures of change', there are five experiments which were carried at IP3. These are described briefly below, with some initial findings. These experiments cover context effects, ambiguous terms and concepts, ambiguous response categories, implicit versus explicit requests for dates of events and proactive dependent interviewing.

i.   Context effects

This experiment was designed to test question order effects in a panel context. A target question was chosen that was likely to be susceptible to context effects: a question about the frequency of behaviour, for which the answer categories used verbal (i.e. vague) frequency labels. A known problem with verbal frequency labels is that different respondents might associate quite different actual frequencies with verbal descriptions such as "always", "sometimes", or "rarely" (e.g. Schaeffer 1991). In addition, how the respondent interprets such frequency labels could be influenced by contextual information. In a panel survey contextual information may change from one wave to the next, such that respondents might interpret the labels differently, and provide different answers to a question, even though their behaviour has in fact not changed.

Respondents were randomly allocated to a split ballot experiment, to test the effects of question context on responses to a question with vague verbal quantifiers. Respondents were allocated to either a low frequency context question ("Since you have been eligible to vote in general elections, how often have you voted? Would you say... always, very often, quite often, sometimes, rarely, never?") or a high frequency context question ("How often do you either listen to the radio or watch TV? Would you say... always, very often, quite often, sometimes, rarely, never?").

All respondents were then asked the same series of target questions:

> "From time to time people discuss political matters or current affairs with other people. Using the categories on the showcard, how often do you talk about politics or current affairs with the following people... Your (husband/wife/partner)? Family members? Friends? Fellow workers? Neighbours? Anyone else, for example, casual acquaintances or strangers?"

The response categories for each of these questions were the same as for the context questions: always, very often, quite often, sometimes, rarely, never.

In wave 4 the experiment was repeated, by swapping the context question for half of the sample.

We expected to see the following effects:

(1) Respondents in the high frequency context group will think of 'always' as meaning 'all the time'. Respondents in the low frequency group will think of 'always' as meaning something less frequent than 'all the time'. In the cross-sectional data we therefore expect the low frequency group to be more likely to say 'always' and less likely to say 'never', than the high frequency group.

(2) In the longitudinal data we expect to see more change in people's responses to the target questions across waves for the group whose context question is swapped, than for the group whose context question remains the same.

Table 5 shows the response distributions for the target questions, separately for the high and low frequency context groups. For the first of the target questions the results are as expected: respondents exposed to the low frequency context were more likely to say

'always' (7.4%), than respondents in the high frequency context group (3.4%) and less likely to say never (11.4%) versus (13.4%). For the following questions about frequency of discussing politics there were no differences between the context groups. Data from wave 4 of the Innovation Panel were not yet available at the time of writing and we can therefore not yet test the longitudinal hypothesis.

Table 5: Context effects (row %)

Talks about politics to...	Context	Always	Very often	Quite often	Sometimes	Rarely	Never	N	P (Chi2)
Husband/wife/partner	low freq	7.4	9.2	19.8	31.8	20.4	11.4	500	
	high freq	3.4	13.0	24.2	29.6	16.4	13.4	537	0.004
Family members	low freq	2.0	7.1	12.0	27.6	26.8	24.4	757	
	high freq	1.8	6.5	14.2	28.7	24.1	24.7	850	0.672
Friends	low freq	1.2	5.3	14.9	30.0	24.2	24.4	757	
	high freq	0.9	6.5	16.0	26.4	26.2	24.0	851	0.557
Fellow workers	low freq	2.7	9.2	13.7	28.7	25.2	20.4	401	
	high freq	1.2	8.5	18.1	26.6	21.0	24.5	481	0.114
Neighbours	low freq	0.7	0.7	2.8	7.7	26.0	62.2	757	
	high freq	0.4	0.2	2.4	9.4	24.0	63.7	851	0.424
Anyone else	low freq	0.3	0.8	2.0	9.2	21.8	65.9	757	
	high freq	0.2	0.5	2.0	11.1	20.7	65.5	850	0.820

Notes: excludes 11 respondents who did not answer the context question.

ii.   Ambiguous terms and concepts

A key purpose of panel surveys is measuring change in people's socio- and economic circumstances over time. Change is however easily overestimated, if respondents answer a given question differently in different interviews, even though their situation has in fact not changed. Spurious change is a known problem in panel surveys, and one of the reasons for using dependent interviewing, where the respondent's answers from a previous interview are used to improve the longitudinal consistency of responses and to prevent response errors that would produce spurious changes (see Jäckle 2009).

This experiment was designed to test whether question wording affects estimates of change. If a survey question is in any way vague, respondents may understand the question differently in one interview than in the next, and as a result may answer the question differently even though their situation is in fact unchanged. We chose several existing BHPS/UKHLS questions which contained terms or concepts for which definitions were potentially ambiguous or where the question wording was in some way complex. Some of the questions had long and complex definitions in the interviewer instructions. For each question we randomly allocated respondents to either the original (potentially ambiguous) question, and to a version in which we tried to remove the ambiguity or make the question easier to understand. The experiment was first carried in wave 2 of the Innovation Panel, and repeated with unchanged allocations to treatments in waves 3 and 4. Treatments were allocated to PSUs, so that each interviewer only worked with one version of the question.

The experimental questions were the following, version A being the original (ambiguous) question, and version B the less ambiguous version.

Disability

> Version A: "Do you have any long-standing physical or mental impairment, illness or disability? By 'long-standing' I mean anything that has troubled you over a period of at least 12 months or that is likely to trouble you over a period of at least 12 months." [Yes/No]
>
> Version B: "Have you been, or are you likely to be, troubled for at least 12 months by any physical or mental impairment, illness or disability?" [Yes/No]

Paid work

> Version A: "Can I just check, did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self-employed?" [Yes/No]
>
> Version B: "Thinking back to the seven days ending last Sunday. During this period, did you do any work for which you receive money or a share of profits?" [Yes/No]
>
> "Did you spend 15 hours or more doing unpaid work in a family business?" [Yes/No]

Savings

> Version A: "Do you save any amount of your income for example by putting something away now and then in a bank, building society, or Post Office account other than to meet regular bills? Please include share purchase schemes, ISA's and Tessa accounts." [Yes/No]

> Version B: "The next questions are about any money you save from your income. Excluding any money you put away to pay for regular monthly or quarterly bills, do you ever put any money away in a bank, building society or Post Office account?" [Yes/No]

> "And do you ever put any money away in share purchase schemes, PEPs, Life Insurance, TESSAs or ISAs?" [Yes/No]

We expected to see the following effect:

With more ambiguous questions, respondents are more likely to answer inconsistently across waves. Therefore we expect the transition rates to be higher than with the less ambiguous version.

Table 6: Transition rates between waves 2 and 3, by question ambiguity (col %)

	Worked		Saved		Disability	
Transition	low ambig.	high ambig.	low ambig.	high ambig.	low ambig.	high ambig.
Yes-Yes	45.9	45.4	43.0	37.2	15.3	26.7
No-No	39.6	43.5	30.7	37.8	62.7	47.3
Total stable	85.5	88.9	73.7	75.0	78.0	74.0
Yes-No	5.7	4.6	11.6	11.8	10.1	10.8
No-Yes	8.8	6.5	14.7	13.2	11.9	15.2
Total change	14.5	11.1	26.3	25.0	22.0	26.0
N	671	632	658	627	622	673
P (Chi2)		0.243		0.046		0.000

The rows in Table 6 present the transition rates in respondent's status between waves 2 and 3. The results are mixed and overall do not provide support for our hypothesis.

We expected the percentage of respondents reporting a change (e.g. moving from work to non-work (Yes-No), or from not having a disability to having a disability (No-Yes)) to be larger with the ambiguous than the less ambiguous version of each question. For disability this does seem to be the case: a larger percentage of respondents report a transition into disability (No-Yes) with the ambiguous version (15.2%) than the less ambiguous question (11.9%). For transitions in work however the opposite is true: more respondents report a change in status with the less ambiguous question (14.5%) than the ambiguous one (11.1%), and for savings there was no difference between question formats.

As a corollary we expected the percentage of respondents reporting the same status in both waves to be higher with the less ambiguous question version. For disability status this is true overall: more respondents report the same status in both waves with the less ambiguous

version (78.0% compared to 74.0%). The overall stability rate however hides differences depending on the status. The overall conclusion is driven by respondents without a disability in both waves. For respondents with a disability the opposite is true: a smaller percentage of respondents report this type of stability with the less ambiguous question (15.3%) than the ambiguous version (26.7%). Similarly for work and savings, fewer respondents reported the same status in both waves with the less ambiguous version.

iii.   Ambiguous response categories

This experiment is a variation on the experiment testing whether ambiguous question wording affects measures of change. In this experiment we tested the impact of ambiguity in question instructions. We selected an existing UKHLS question for which the instructions were potentially ambiguous. The question about current labour market activity is a closed question where the respondent is asked:

> "Which of these **best** describes your current employment situation? Self employed / In paid employment (full or part-time) / Unemployed / Retired / On maternity leave / Looking after family or home / Full-time student / Long-term sick or disabled / On a government training scheme / Unpaid worker in family business / Doing something else".

The question instructions are potentially ambiguous in that the response categories are not mutually exclusive, for example a person can be retired but also have a job, and there are no instructions about how to select the main category if more than one applies. This ambiguity means that respondents to whom more than one status applies may report one status in one interview, and the other in the next. This would suggest that their situation has changed, although in fact it has not, again leading to an over-estimation of change.

The experiment randomly allocated respondents to either the original 'forced choice' version of the question, or a version where respondents were asked to 'tick all that apply': "Which of these describe your current employment situation?" Respondents who reported more than one activity were then asked a follow-up question: "And which would you consider your main current employment situation?" The treatment allocation was at the PSU level.

At wave 2 the experiment was crossed with a showcard experiment. For both question versions half the sample received the question with a showcard, the other half without a showcard. The showcard treatment was allocated at the PSU level. At wave 3 all respondents received a showcard.

We expected to see the following effect:

Allowing for multiple responses, the proportion of the sample reporting the same labour market activity in both waves is likely to be higher with the 'tick all that apply' format, than the 'forced choice' format.

Table 7: transition in current labour market status, by question format

		Forced choice (col %)	Tick all that apply (col %)
W2 showcard	Transition	13.8	15.3
	Same status	86.2	84.7
	N	311	300
	P (Chi2)		0.598
Full sample	Transition	14.6	12.9
	Same status	85.4	87.1
	N	639	605
	P (Chi2)		0.395

Since half the sample did not have a showcard in wave 2, while showcards were used for all in wave 3, Table 7 presents the results for the wave 2 showcard sample only, and for the combined showcard and no showcard wave 2 sample.

The results do not support our hypothesis. There were no significant differences in the proportion of respondents reporting a change in labour market status between waves. Future work could examine transitions in different types of activities, and between employment, unemployment and inactivity.

iv.   Implicit versus explicit requests for dates of events

This experiment is a further variation on the theme that the ambiguity of questions or instructions might affect the measurement of change. The way a question is worded may affect which recall and judgment strategies respondents use in order to compute an answer. This experiment tested whether the way in which questions about the dates of events are formulated affects the quality of date information. We contrasted two versions. In the first version respondents were explicitly asked to provide a date: "In which month and year did you...?". This wording made it clear to respondents that an exact date was required. In the second version the request was ambiguous: "How long have you...?" Respondents could answer by either providing the date or the elapsed duration since the event. We considered this version to be an implicit request for a date. We implemented the experiment using several items:

Residential moves
        Explicit request: "In what month and year did you move to **this** address?"
        Implicit request: "How long have you lived at **this** address?"

Respondents interviewed previously were asked about moves since the previous interview. New household members who were interviewed for the first time were asked about the date they moved to their current address.

14

Health conditions
>   Explicit request: "In which year were you first told you had [Health Condition]?"
>   Implicit request: "When were you first told you had [Health Condition]?"

Private pension schemes
>   Explicit request: "In which year did you join this pension scheme?"
>   Implicit request: "How long have you been a member of this pension scheme?"

These experiments were first implemented in wave 2 of the Innovation Panel, and repeated at waves 3 and 4. In waves 3 and 4 each date question was followed by a closed question about how the respondent had come up with the answer: "How did you come up with that date? Guessed / Knew the exact date / Related it to their age / Related it to the date of another event / Remembered something about the event that suggested when it happened / Doesn't know how they came up with the date".

We expected to see the following effects:

(1) The quality of date information is likely to be better with explicit than implicit requests. As a result we expect the elapsed duration since the event to be reported with more 'heaping' when the date request is implicit. That is, we expect more respondents to report durations (or dates that imply durations) of multiples of 12 months.

(2) We expect the differences in data quality to be due to differences in the way respondents process the question, and in particular in the recall strategy used to retrieve/compute the date. With the explicit request we expect more respondents to directly recall the date, or information related to the date. With the implicit request we expect more respondents to guess or estimate.

For the question about the date of joining a pension scheme the results are as expected. Figure 1 shows that with the implicit date request there is heaping in the reported durations at multiples of 5 years. 12% if respondents reported a duration of 10 years, 15% a duration of 20 years. In total 48% of respondents report a duration of either 5, 10, 15, 20 or 25 years.

The distribution of durations obtained from the explicit date questions is much smoother. At most 6% of respondents reported the same duration (3, 10, 22 and 25 years). 22% reported a duration of 5, 10, 15, 20 or 25 years.

For the questions about the dates of moving and the dates when health conditions were first diagnosed the sample sizes are very small and no clear patterns emerge.

Figure 1: Duration since joining pension scheme

Graphs by type of date Q
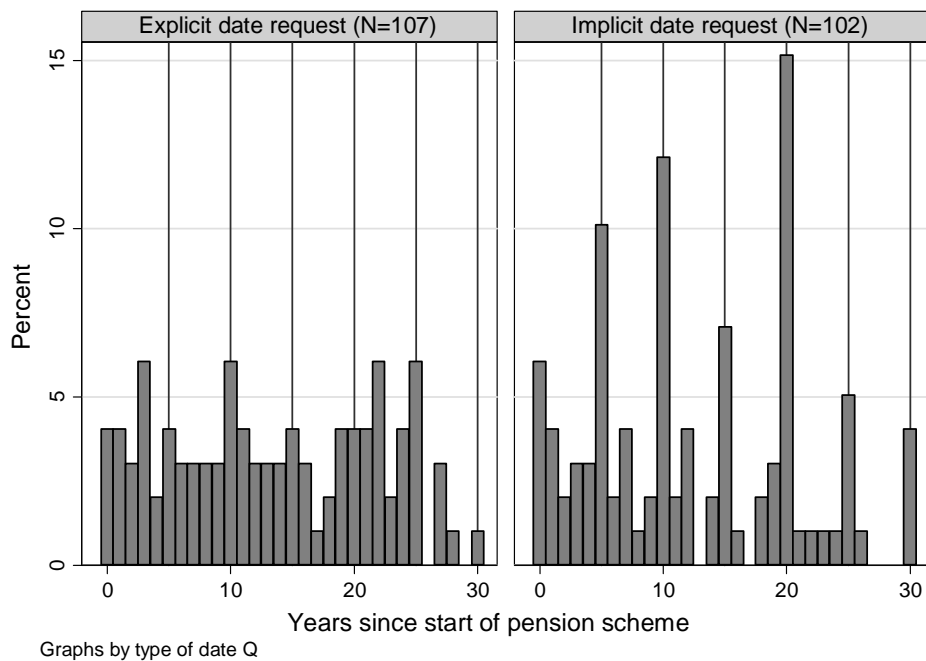
Because the numbers of respondents who reported dates of events (moving house, joining pension scheme, onset of health conditions) were small, we pooled the questions about the response strategy for all events. This means that Table 8 includes multiple responses from some respondents: 88% of respondents included in the table provided only one date; 9% provided two dates; 3% provided three or more dates. We had designed the response strategy questions as 'tick all that apply'. For two questions this was implemented correctly, for two questions it was implemented as forced choice. Since few respondents in the tick all that apply format selected more than 1 response option, we recoded the variables to include only the first mention.

The results summarized in Table 8 suggest some support for our hypothesis that the wording of the request affects which strategies respondents use to compute an answer. With the explicit request respondents were more likely to recall the date or relate it to the date of another event or characteristics of the event. With the implicit date request respondents were more likely to guess the date or relate it to their own age.

Table 8: Response strategies for dates of events, by wording of date question

| | Explicit date request | | Implicit date request | |
	col %	N	col %	N
Guessed	23.8	51	34.9	67
Knew the exact date	44.9	96	35.9	69
Related it to their age	7.0	15	13.5	26
Related to date of other event	15.4	33	12.0	23
Related to characteristics of event	6.5	14	2.6	5
Doesn't know	2.3	5	1.0	2
Total		214		192

Future work will examine respondents' choices of how to answer the implicit question. For the date of joining a pension scheme 96 respondents reported a duration, only 7 reported a date. Future work will also combine the experimental data from waves 2, 3 and 4 to examine the effects in the questions on residential moves and health conditions, for which the sample sizes in each individual wave are very small.

v.   Proactive dependent interviewing

With dependent interviewing answers given by respondents in a previous interview are fed forward and incorporated into the questionnaire script, to verify that reported changes in the respondent's situation are in fact true and not the result of a reporting error. With proactive dependent interviewing (PDI), the fed forward information is incorporated into the question text. The previous information can be used in different ways (see Jäckle 2009): (1) respondent can be asked to confirm the previous information (Remind, confirm), (2) the previous information can simply be used as a starting point to ask about the current situation or events since the previous interview (Remind, continue), or (3) respondents can be asked whether the previous report still applies (Remind, still). With the 'Remind, still' format, there is concern that respondents may simply say 'yes' to the information presented to them. The tendency for some respondents to answer 'yes', regardless of whether the correct answer to the question given their situation demands a 'yes' answer, could be for reasons of satisficing (Krosnick 1991) or cooperative linguistic engagement (Raymond 2003). That is, the 'Remind, still' question design may lead to change being underestimated if respondents falsely confirm a previous status as still applying to them.

This experiment was designed to test whether the wording of 'Remind, still' questions affects the measurement of change. Respondents were randomly allocated to two versions of PDI questions. In one version respondents were reminded of their answer from the previous interview and asked whether this was still the case. In the other version respondents were instead asked whether this had changed. The treatments were allocated at the level of the PSU. The experiment was implemented using four different questions:

General health

> Remind, still: "The last time we interviewed you on <date of interview> , you said that, in general, your health was <[excellent]/[very good]/[good]/[fair]/ [poor]>. Is that still the case?"

> Remind, changed: "The last time we interviewed you on <date of interview> , you said that, in general, your health was <[excellent]/[very good]/[good]/[fair]/ [poor]>. Has that changed?"

Whether job is permanent

> Remind, still: "Last time we interviewed you on <date of interview>, you said that leaving aside your own personal intentions and circumstances, your job was <[a permanent job]/ [not a permanent job in some way]>. Is this still the case?"

> Remind, changed: "Last time we interviewed you on <date of interview>, you said that leaving aside your own personal intentions and circumstances, your job was <[a permanent job]/ [not a permanent job in some way]>. Has this changed?"

Employed working hours

> Remind, still: "Last time we interviewed you, you said that in your (main) job, you were expected to work <number> hours in a normal week, excluding overtime and meal breaks. Is this still the case?"

> Remind, changed: "Last time we interviewed you, you said that in your (main) job, you were expected to work <number> hours in a normal week, excluding overtime and meal breaks. Has this changed?"

Self-employed working hours

> Remind, still: "Last time we interviewed you, you said that you usually work <number> hours in total each week in your job. Is this still the case?"

> Remind, changed: "Last time we interviewed you, you said that you usually work <number> hours in total each week in your job. Has this changed?"

We expected to see the following effect:

If there is a tendency to agree (either because of satisficing or because of social desirability/ conversational norms), then the percentage of respondents saying 'yes' their status is still the same will be higher than the percentage of respondents saying 'no' their status has not changed – and vice versa.

Table 9: Change in status by PDI wording

| Survey item | PDI treatment | | | | |
| | Remind, still? | | Remind, changed? | | |
	col %	N	col %	N	P (Chi2)
General health is the same	88.8	690	75.1	562	
General health has changed	11.2	87	24.9	186	0.000
Permanency of job same	95.3	386	77.1	276	
Permanency of job changed	4.7	19	22.9	82	0.000
Working hours same – emp	80.0	268	64.8	206	
Working hours changed – emp	20.0	67	35.2	112	0.000
Working hours same – semp	73.1	38	62.1	18	
Working hours changed – semp	26.9	14	37.9	11	0.304

Notes: emp=employees, semp=self-employed

The results summarised in Table 9 provide support for our hypothesis. The percentage of respondents who said 'yes' their status was still the same was between 11 and 18 percentage points higher than the percentage of respondents who said 'no' their status had not changed.

Future work will examine response latencies to for 'yes' and 'no' responses in both treatment groups. In both versions of the DI question respondents who indicated a change were asked a follow-up question about what their current situation was. Future work will also check whether the answers to the follow-up questions confirm the change in circumstances reported in the DI question.

c.   Branched versus un-branched questions

This experiment was designed to test for differences in measurement between branched and un-branched attitude scales. The background to branching design research was provided by Armstrong, Denniston and Gordon (1975). They use the decomposition principle, which shows that by breaking a decision task up into component decision parts, the accuracy of the final decision is increased. From this, one could imagine that breaking an attitudinal item into its component parts would increase the accuracy of the final report. In practice, this is applied by firstly asking the respondent about the direction of their attitude, and then using a follow up question to measure the intensity of the attitude (Krosnick and Berent, 1993).

The aim of this experiment was to compare and evaluate branched and unbranched scales for the measurement of attitudes in longitudinal surveys. Specifically, we aim to establish which response scale format provides the most reliable and valid responses for attitude measurement. We also intend to establish the underlying mechanisms involved in different formats, and therefore reasons for differences in responses between formats.

The research questions considered here are:
        1. Are there differences in responses between branched and unbranched questions?
        2. Which format provides the most reliable responses?

A split ballot experiment was used, with half of the respondents receiving branched versions of questions and the other half receiving unbranched versions of the same questions. A five point scale was chosen due to the difficulty in presenting a fully labelled, seven point scale to a respondent without a showcard. Two batteries of questions were used, concerning political efficacy and neighbourhood social cohesion.

Response distributions differ depending on the branching condition: those in the branched treatment condition are more likely to use the extreme response options (i.e., strongly agree and strongly disagree) compared to those in the unbranched group.

Mean scores for each individual for each battery of questions reflect differences between branched and unbranched questions in terms of response distributions. Those in the branched treatment group are significantly more likely to have an extreme mean score. Conversely, those in the unbranched condition are more likely to have a mid-range score.

For each question we then derived an indicator of whether the respondent had selected an extreme response option, versus one of the middle options. The extreme score was regressed against the branching condition, the results of which are shown in table 10. For both batteries of questions, respondents in the branched group were significantly more likely to use extreme response options compared with those in the unbranched treatment condition.

Table 10: Likelihood of choosing an extreme option

	Political efficacy extreme score		Neighbourhood extreme score	
	B	S.E.	B	S.E.
Intercept	1.136***	0.038	1.132***	0.037
Unbranched	-0.118***	0.024	-0.188***	0.024
R^2	0.0140		0.0378	
N	1558		1577	

* significant at p<0.05 ** significant at p<0.01 *** significant at p<0.001

To examine the reliability of the branched and unbranched scales, we calculated Cronbach's Alpha scores for each of the two scales, separately for branched and unbranched questions.

Table 11: Cronbach's Alpha scores for political efficacy and neighbourhood scales

	Average interim covariance	Number of scale items	Scale reliability coefficient
Political efficacy branched	0.4147	4	0.6082
Political efficacy unbranched	0.3121	4	0.6067
Neighbourhood branched	0.5483	4	0.7648
Neighbourhood unbranched	0.4038	4	0.7933

The results of Cronbach's Alpha (table 11) show that there appears to be very little difference in terms of reliability between the branched and unbranched scales - the scale reliability coefficients are very similar for each pair of scales.

To conclude, the data show evidence of response differences between branched and unbranched scales, particularly the higher rate of extreme responding in the branched format. Looking at which scale type is more reliable, Cronbach's Alpha seems to show the branched and unbranched formats are equally reliable.

Further work will examine the effect of branching format on reliability using Confirmatory Factor Analysis and Graded Response Models. Further to this, correlations of criterion variables will be examined to test the validity of the branched and unbranched scales. Timing data will also be used to look at any differences in administering these scales, with an interest particularly in whether any gains in reliability using branched questions are offset by the potential additional costs incurred because of the time taken to administer the questions.

d.   Improving measures of wealth

One of the areas of new content for the UKHLS at Wave 4 is the collection of wealth, assets and debt information. This information was collected on the BHPS at Waves 5, 10 and 15. However, the UKHLS gives us an opportunity to develop and improve on the questions used on the BHPS. Questions around the area of wealth and debt are potentially sensitive and may be cognitively difficult for respondents to answer. We reviewed how these topics are approached in other longitudinal surveys. On IP3 we tested four different question designs for collecting the amount of money held in savings and investments. By comparing the data resulting from each design to comparable measures from the UK Wealth and Assets Survey, we aim to identify the most effective way of asking these questions.

In each design an initial question asked respondents to report which specific savings or investments they held. Then, four different approaches to obtaining the amount held resulting in a cross of two separate treatments: (a) whether to ask for item-by-item amounts held in savings or investments ("itemised") versus obtaining only aggregate sums for each of savings or investments; and (b) whether to obtain information on all forms of savings or investments from each responding adult ("individual reporting") versus targeting the household respondent ("financial reporting") for information on certain types of investments. Households were randomly assigned to one of the four experimental treatments. All individuals interviewed as part of the household were subjected to the group's treatment.

Table 12. Schematic design of IP3 experimental treatments for wealth questions

	Individual Reporting	Financial Reporting
Aggregate Amounts	*Group 1*	*Group 2*
Itemised Amounts	*Group 3*	*Group 4*

The aim of this analysis was to compare these four designs in terms of the missing data generated and the total amounts obtained for households. We examine these values in light of similar amounts obtained from BHPS Waves 10 and 15 which share the Group 1 design.

We also compare amounts obtained through each design to amounts computed from the UK Wealth and Assets Survey (WAS) which we treat as a "gold standard" given its design.[5]

Table 13, below, sets out the mean and median household amounts for each treatment groups, along with the figures from the BHPS and the WAS for comparison. The table also reports on the percentage of missing data. Those individuals who did not select a particular savings or investment type are assumed to have no money in that financial instrument (treated as zero). Those who refused or said that they didn't know have been dropped from the analysis. The table gives the results for savings and investments together, savings alone and investments alone.

Regarding the total amount held in savings and investments at the household level, there are significant differences across experimental treatments. Experimental treatments involving a financial reporter seem to generate mean and median values closer to the Wealth and Assets Survey as compared to those based on individual reports. The ratios for Group 2 and Group 4 are close to one, suggesting that the quality of data collected in the IP is comparable, though there is less variation in the IP than in the WAS as is to be expected given its small sample size. The picture is a little different if we focus the analysis on savings only, when itemised reporting seems to yield means which are closer to the WAS values. If we focus on investments only, we find the opposite; the role of a financial reporter – rather than collecting the information from all adults – is important. Using a financial reporter significantly also reduces the amount of missing data on investments. It may be, then, that investments are something which are relatively rare, held at the household-level and the details are likely to be known by just one person in the household, whilst savings are more common and familiar and may be held at the individual-level and so are multiple people in the household are able to respond.

An analysis of variance exploring the main effects of using a financial reporter and itemisation on mean amounts reported suggests that for total savings and investments combined, there is no main effect of itemisation in explaining variability of means whereas the use of a financial reporter seems to capture most variability. This pattern is mirrored in the results for investments alone where we observe a significant main effect of using a financial reporter but no effect of itemisation and no interaction between itemisation and a financial reporter. For savings, itemisation does play a role whereas using a financial reporter has no effect and there is no interaction between these designs.

---

[5] All data have been inflated using the consumer price index to 2010 values for comparison. Moreover, the WAS does not collect financial details of full-time students between the ages of 16 and 18, therefore this segment of both the BHPS and IP samples was excluded from analysis. There are some caveats to this comparison. First, the categories over which WAS respondents report are slightly different to the categories used in both the BHPS and the IP. The main differences are that amounts in National Savings accounts are combined with other savings and deposit accounts, and that amounts in premium bonds, National Savings bonds and National Savings certificates are collected as a single category rather than three separate categories. Secondly, cross-sectional response weights were used to obtain amounts in the WAS and the BHPS whereas the IP figures are obtained from unweighted data. This IP3 analysis is still preliminary and weights have not yet been computed. Differences between numbers may be due, therefore, to sample composition differences rather than overall design.

Using paradata from the individual interview, we find that there is negligible difference in respondent burden, as measured by administration time, across the experimental groups. Within the financial reporter treatments, itemisation does not seem to make any difference to either mean or median administration times.

Based on these results, we believe that the design of Group 4 provides consistently better data with little difference in administration burden over the current BHPS design. The amount of missing data obtained about investments is negligible as compared to asking all respondents for investment amounts. Moreover, the mean and median values for this group are consistently closer to the WAS data than other groups suggesting that the data obtained using this design tends to be more accurate. For these reasons, we believe this design would be better suited for inclusion in Wave 4 of *Understanding Society.*

Table 13. Total household savings and investments, comparing the four IP3 experimental treatments to the BHPS Waves 10 and 15 and the Wealth and Assets Survey.

	Total household savings and investments			Household savings			Household investments		
	Mean	Median	Missing	Mean	Median	Missing	*Mean*	*Median*	Missing
BHPS W10	£22,440	£5,534	17.8%	£11,355	£3,690	17.4%	£19,654	£3,690	16.1%
BHPS W15	£25,670	£6,764	21.4%	£15,554	£4,919	20.8%	£19,933	£2,690	19.6%
WAS	£45,657	£8,202	11.2%	£24,357	£6,026	8.2%	£43,137	£6,999	10.7%
Group 1: Agg, Ind	£18,616	£6,000	23.9%	£12,720	£5,000	21.1%	£12,142	£2,000	14.4%
ratio to WAS	*0.41*	*0.73*		*0.52*	*0.83*		0.28	0.29	
Group 2: Agg, Fin	£46,321	£10,250	26.5%	£12,902	£5,750	24.2%	£61,650	£20,000	5.4%
ratio to WAS	*1.01*	*1.25*		*0.53*	*0.95*		1.43	2.86	
Group 3: Item, Ind	£22,579	£6,050	24.4%	£16,943	£6,000	24.7%	£15,896	£3,000	18.9%
ratio to WAS	*0.49*	*0.74*		*0.70*	*1.00*		0.37	0.43	
Group 4: Item, Fin	£52,580	£10,000	29.0%	£22,964	£5,000	25.0%	£48,692	£6,000	4.2%
ratio to WAS	*1.15*	*1.22*		*0.94*	*0.83*		*1.13*	*0.86*	

Notes: Respondents for whom an item was coded as inapplicable were treated as having value zero. Item missings due to "refusal" or "don't know" were dropped from the analysis.

e.   Incentives and response

As at IP1 and IP2, IP3 included an incentive experiment. One of the groups at IP2, which had received £10 at both IP1 and IP2, was split into two groups, with one still receiving £10 and the other group reduced to £5. This was done to test the effect of reducing the level of incentive at the third wave rather than the second wave. This manipulation gives us six incentive groups. Table 14 below sets out the six groups. Three groups (A, B, C) have had the same level of incentive since IP1; receiving either £5 each, £10 each or £5 each with the promise that this would increase to £10 each if all eligible adults participated ("£5-to-£10"). Group D started of as £5-to-£10 at IP1 but was reduced to £5 at IP2 and remained at that level for IP3. Group E started off at £10 at IP1, was reduced to £5 at IP2 and remained at that level for IP3. The comparison of interest is between this group and Group F, where the reduction from £10 to £5 occurs at IP3.

Table 14: Incentive groups across the Innovation Panel waves 1 to 3.

Group	IP1	IP2	IP3
A	£5	£5	£5
B	£10	£10	£10
C	£5-to-£10	£5-to-£10	£5-to-£10
D	£5-to-£10	£5	£5
E	£10	£5	£5
F	£10	£10	£5

The incentive was pre-paid and sent to the sample member with the advance letter. All adults in the household received an advance letter with the incentive, including those who had turned 16 since the previous interview. The letter was tailored according to whether the individual had participated in the previous wave, had not participated or was still aged 15. This means that those who did not participate in the previous wave still received an unconditional voucher. Everyone in the household was allocated to the same incentive group, and new entrants to the household and split-offs from the household are treated in the same way as the original household.

The analysis below treats untraced movers as non-contacts. Table 15 shows the household response rate by incentive group. Compared to the group who received £5 at every wave (A), those households who always received £10 (B), were significantly more likely to be productive ($p<0.1$). Households who had received £5-to-£10 at IP1 and then £5 only in IP2 and IP3 (D), were significantly more likely to refuse than those households who had always received £5 (A). There was no significant difference between the group who had received £10 for one year (IP1) before the incentive was reduced to £5 (E), and those who had received £10 for two years (IP1 and IP2) before the incentive was reduced (F).

Table 15: Household response by incentive group

	£5 always %	£10 always %	£5-to-£10 always %	£5 (£5-to-£10 at IP1) %	£5 (£10 at IP1) %	£5 (£10 at IP1/2) %	All %
Productive	64.8	73.7*	68.6	60.6	66.9	67.7	66.1
No contact	11.8	8.3	11.2	9.7	9.4	13.8	10.8
Refusal	16.9	12.0	17.1	24.3**	17.3	13.1	17.5
Other non-response	6.5	6.0	3.1*	5.4	6.4	5.4	5.6
n	492	133	258	259	266	130	1,538

* p<0.1 ** p<0.05. Reference group is those who received £5 at every wave (A)

Table 16 shows the response for individuals within productive households. Compared to individuals who had always received £5, those who had initially received £10 (at IP1), reduced for IP2 and IP3 were more likely to give a full interview. Individuals in all other incentive groups were less likely to refuse, compared to those who had always received £5.

Table 16: Individual response within productive households by incentive group

	£5 always %	£10 always %	£5-to-£10 always %	£5 (£5-to-£10 at IP1) %	£5 (£10 at IP1) %	£5 (£10 at IP1/2) %	All %
Full interview	79.9	81.7	81.2	81.8	85.1**	82.6	81.7
Proxy interview	5.9	6.9	6.7	8.2	5.9	10.3*	6.8
No contact	2.3	1.5	1.2	2.8	3.1	1.9	2.2
Refusal	9.6	4.5**	6.4*	6.0*	4.5**	4.5**	6.6
Other non-response	2.3	5.4**	4.6*	1.3	1.4	0.6	2.5
n	612	202	329	318	356	155	1,972

* p<0.1 ** p<0.05. Reference group is those who received £5 at every wave (A)

Individuals within productive households who had received two waves of £10 incentives before being reduced to £5 were more likely to give a proxy interview than those who had only one wave of £10 before the reduction (p<0.1). There does not seem to be an effect of the £5-to-£10 group of increasing within-household response rate. The individual response within productive households is not significantly different for this group than for those households who had always received £5 each.

To conclude, at the household level only the £10 incentive group was marginally more likely to be productive than the £5 group. For individuals within productive households, the £10 group had fewer refusals than the £5 group but more other types of non-response. Reducing the level of incentive after the first wave of a longitudinal survey does not seem to have had a significant effect on household response. For individuals, it appears that if a reduction in the level of incentive must happen, to do it early (after one wave) is better than later (after two waves). There is still further work to do in this area, including looking at whether increasing the incentive once a panel has been established has an effect on response.

5   Implications for survey design in the future

The Innovation Panel has been established as an important source of experimental data for methodological research. Over the first three waves it has looked at a number of areas within survey and questionnaire design. In terms of survey design, the IP has carried experiments with mixed modes, different incentive strategies, the use of showcards and different types of advance materials. Within the questionnaire, the IP has carried experiments with different ways of eliciting measures of household and personal consumption, the number of points on a response scale, the extent of labelling on a response scale, panel conditioning, the effect of question wording of measures of change over time, branched versus un-branched questions and different ways of obtaining measures of household wealth. With IP3, we have opened up the survey to researchers from around the world who are interested in experimenting with ways to improve survey measurement. We will do the same for IP4 and IP5.

In addition to being a major resource for methods researchers, the Innovation Panel is also an important part of the wider UKHLS. The lessons learnt from the IP feed into the design for the main-stage of the UKHLS. IP3 was instrumental in helping us decide how to collect the complex and potentially sensitive information about the wealth and assets of a household. For Wave 4, the financial instruments relating to investments will be separated from those relating to savings. The investments will be asked at the household level, whilst savings will be asked at the individual level. Cognitive testing of these questions, carried out by the National Centre for Social Research (NatCen), suggested a more respondent-friendly way to name and group the different financial instruments which make it easier for people to respond.

Measures of cognitive function were carried at Wave 3. We viewed the cognitive function measures at IP3 as a test of our ability to carry and administer these types of measures. The measures carried at IP3 were well-known and required little operationalisation to implement on the survey. There were some changes though, following the trial on IP3 and cognitive interviewing. The FAS test was dropped and was replaced by animal naming where the respondents had to name as many animals as possible in one minute. The reason for this was that during cognitive interviewing it became apparent that there was some problems with respondents understanding the instructions about proper nouns and some phenomic confusion with, particularly, the with the letter 'A'. The prospective memory test was not carried at Wave 3. During the cognitive interviewing phase, we found that the test did not work effectively with non-native English speakers who did not recognise what was meant by "initials". Additional cognitive function tests were carried on the main-stage of Wave 3; the immediate and delayed recall test and the number series test. These measures are standard and were tested using cognitive interviewing techniques. For more information on the cognitive interviewing see Gray et al, 2011.[6]

---

[6] http://research.understandingsociety.org.uk/publications/working-paper/2011-03.pdf

6   Plans for IP4

IP4 went into the field in March 2011 and was fully face-to-face. Those households who were not interviewed at either IP2 or IP3 were reviewed and any adamant refusals removed from the issued sample at IP4. In addition, there was a refreshment sample recruited and interviewed at IP4, with the aim of an additional 500 achieved households. A number of experiments carried thus far on the IP are longitudinal in design and so were carried at IP4. A competition for experiments was held in May-June of 2010 and there were 24 proposals submitted and one which had been carried over from the IP3 competition. A summary of the experiments carried on IP4 are below. There will be more information on these in a future Working Paper.

    i.   Improving the reliability of measures of change from panel surveys. This is a continuation of an experiment carried at IP2 and IP3. At IP4 a random allocation of households will get the same type of questions as previously whilst the rest will have a change of question style.

    ii.   Panel conditioning in measurement. This is a continuation of an experiment started at IP1 using questions on height and weight, and environmental behaviour.

    iii.   A comparison of branched versus un-branched rating scales for the measurement of attitudes in panel surveys. This is a continuation of an experiment carried at IP3 and described in section 4c above.

    iv.   Mode effects in a self-completion instrument. At Wave 3 of the main-stage of the UKHLS, the adult self-completion was a CASI instrument, after being a paper self-completion in the first two waves. In this experiment adults in a randomly allocated half of households will receive the CASI instrument, while those in the other half will receive the paper self-completion. The aim is to look at the effects on response rates for this instrument and measurement differences between the two modes.

    v.   Context effects. Experimenting with the placement of a question about fertility decisions.

    vi.   Incentives. A continuation of the incentive experiments.

    vii.   Context and question wording on request for consent to data linkage. This experiment manipulates the placement of a request for consent to link survey data to administrative benefit records. It also looks at the effect of asking for consent independently or in a proactive dependent way.

    viii.   Inter-wave mailing experiment to increase registration on the Participants website. This uses the between-wave mailing to encourage sample members to register with the Participant website. It allocates sample members to either a paper or email mailing, with an incentive to register or no incentive.

    ix.   'Early Bird' experiment. This encourages sample members to contact the interviewer to arrange an appointment for their interview in the run-up to the start of fieldwork. The sample is split into three groups; Early Bird with an additional incentive, Early Bird with no additional incentive and no Early Bird.

    x.   Interviewer observations. This experiment examines the utility of observation data obtained for the IP4 refreshment sample cases.

    xi.   The effect of additional information during the re-issue phase. Half of those households that are re-issued after the main fieldwork period will receive an information leaflet seeking to respond to the main reasons people give for refusal. The measure of interest is the conversion rate of the re-issued households.

xii. Mode preference. This is a two-wave experiment for IP4 and IP5. Respondents at IP4 will be asked for their preferred mode of interview. At IP5 the mode will be randomly allocated, and the combination of preferred mode and actual response can be used to predict likely response in different modes.

A number of these experiments were audio-recorded for additional analysis, with the consent of the respondent. The IP4 questionnaire is available on the *Understanding Society* website at: http://data.understandingsociety.org.uk/questionnaires/wave-4.

References

Armstrong, J.S., Denniston, W.B. and Gordon, M.M. (1975) "The Use of Decomposition Principle in Making Judgements": *Organizational Behavior and Human Performance*, 14: 257-263.

Burton, J., Laurie, H. and Uhrig, SC N. (eds.) (2008) "Understanding Society. Some preliminary results from the Wave 1 Innovation Panel". Understanding Society Working Paper 2008-03. Colchester: University of Essex.
http://research.understandingsociety.org.uk/publications/working-paper/2008-03

Burton, J., Laurie, H. and Uhrig, SC N. (eds.) (2010) "Understanding Society Innovation Panel Wave 2: Results from Methodological Experiments". Understanding Society Working Paper 2010-04. Colchester: University of Essex.
http://research.understandingsociety.org.uk/publications/working-paper/2010-04

Gray, M., D'Ardenne, J., Balarajan, M. and Uhrig, SC N. (2011) "Cognitive Testing of Wave 3 Understanding Society questions". Understanding Society Working Paper 2011-03. Colchester: University of Essex.
http://research.understandingsociety.org.uk/publications/working-paper/2011-03

Jäckle, A. (2009) "Dependent Interviewing: A Framework and Application to Current Research". Pp. 93-111 in *Methodology of Longitudinal Surveys*, edited by P. Lynn. Chichester: Wiley.

Krosnick, . A. (1991) "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys". *Applied Cognitive Psychology* 5:213-236.

Krosnick, J.A. and Berent, M.K. (1993) "Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format". *American Journal of Political Science*, 37(3): 941-964.

Lynn, P. (2009) "Sample Design for Understanding Society". Understanding Society Working Paper 2009-01. Colchester: University of Essex.
http://research.understandingsociety.org.uk/publications/working-paper/2009-01

Raymond, G. (2003) "Grammar and Social Organisation: Yes / No Interrogatives and the Structure of Responding" *American Sociological Review* 68:939-967.

Schaeffer, N.C. (1991)"Hardly Ever or Constantly? Group Comparisons using Vague Quantifiers". *Public Opinion Quarterly* 55:395-423.