Understanding Society

Working Paper Series

No. 2012 – 06

September, 2012

Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd[1], Emily Gilbert[1], Jonathan Burton[1], Annette Jäckle[1], Olena Kaminska[1], SC Noah Uhrig[1], Matt Brown[2] and Lisa Calderwood[2]

[1] Institute for Social and Economic Research, University of Essex

[2] Centre for Longitudinal Studies, Institute of Education

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

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

Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd[1], Emily Gilbert[1], Jonathan Burton[1], Annette Jäckle[1], Olena Kaminska[1], SC Noah Uhrig[1], Matt Brown[2] and Lisa Calderwood[2]

[1] Institute for Social and Economic Research, University of Essex
[2] Centre for Longitudinal Studies, Institute of Education

Abstract

This paper presents some preliminary findings from the Wave 4 Innovation Panel (IP4) of *Understanding Society*: The UK Household Longitudinal Study. *Understanding Society* is a major new panel survey for the UK. In March 2011, the fourth wave of the Innovation Panel was fielded. This paper describes the design of IP4, the experiments carried and the preliminary findings from early analysis of the data.

Key words: longitudinal, survey methodology, experimental design, respondent incentives, questionnaire design, branching, dependent interviewing.

JEL classification: C80, C81, C83

Contact: Jonathan Burton (jburton@essex.ac.uk) Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments

Jonathan Burton (ed.)

Contributors: Sarah Budd[1], Emily Gilbert[1], Jonathan Burton[1], Annette Jäckle[1], Olena Kaminska[1], SC Noah Uhrig[1], Matt Brown[2] and Lisa Calderwood[2]

[1] Institute for Social and Economic Research, University of Essex
[2] Centre for Longitudinal Studies, Institute of Education

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 4 of the Innovation Panel in the spring of 2011. The methodological tests included an experiment testing the effects of different incentives offered to respondents in advance of fieldwork on response rates, the use of an "Early Bird" request for sample members to contact their interviewer to arrange an appointment for an interview, the collection of interviewer observations, and the use of different re-issue letters. Further tests examined different ways of designing survey questions, including branched versus unbranched response categories, asking about mode preference, (re-) asking consent to data linkage, context effects, implicit versus explicit requests for dates, panel conditioning and mode effects in a self-completion instrument.

Contents

1 Introduction 1

2 *Understanding Society*: the UKHLS 1

3 Innovation Panel Wave 4: Design 2

a. Call for experiments 3

b. Sample 3

c. Interview schedule 4

d. Response rates 4

4 Experimentation in IP4 6

a. Branched versus un-branched questions (Emily Gilbert) 6

b. Re-issue Experiment (Sarah Budd) 7

c. Want to be an Early Bird? Can encouraging respondents to contact interviewers to make appointments increase co-operation and save costs? (Matt Brown and Lisa Calderwood) 8

d. Asking Respondents about Mode Preference (Olena Kaminska) 10

e. Interviewer Observations (Olena Kaminska) 11

f. Incentives (Jonathan Burton) 14

g. Consent to data linkage (Jonathan Burton) 19

h. IP4 Context effects (Noah Uhrig) 24

i. Panel Conditioning in IP4 (Noah Uhrig) 30

j. Between-wave mailing experiment (Jonathan Burton) 33

k. Mode effects in a self-completion instrument (Jonathan Burton) 35

l. Implicit versus explicit requests for dates of events (Annette Jäckle) 40

m. Ambiguous terms and concepts (Annette Jäckle) 44

n. Wording of dependent interviewing questions (Annette Jäckle) 44

5 Plans for IP5 45

REFERENCES 47

1 Introduction

This paper presents early findings from the fourth wave of the Innovation Panel (IP4) of *Understanding Society*: The UK Household Longitudinal Study (UKHLS). *Understanding Society* is a major panel survey for the UK. The first two waves of data collection have been completed (January 2009 to March 2011) and the full Wave 1 and the Wave 2 Year 1 data are available from the UK Data Archive, with full Wave 2 data later in 2012. Data for the first four waves of the Innovation Panel are also available from the Data Archive[1].

One of the features of *Understanding Society*, 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 *Understanding Society* since it was first proposed. Part of this drive for innovation is embodied within the Innovation Panel (IP). This panel of almost 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 *Understanding Society*. 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 and the third wave (IP3) in April-June 2010. Working Papers which cover the experimentation carried out in all three innovation panels are available from the *Understanding Society* website.[2] The data from the first four waves of the innovation panel are held at the UK Data Archive and are made available via the Economic and Social Data Service.[3] In March 2011, IP4 was fielded. This paper describes the design of IP4, the experiments carried and some preliminary findings from early analysis of the data. Section 2 outlines the main design features of *Understanding Society*. Section 3 describes the design and conduct of IP4. Section 4 then reports on the experiments carried at IP4. Section 5 concludes the paper by outlining plans for IP5.

2 *Understanding Society*: the UKHLS

Understanding Society 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 NatCen Social Research (NatCen). *Understanding Society* 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. *Understanding Society* provides high quality,

[1] http://www.esds.ac.uk/findingData/ukhlsTitles.asp
[2] http://research.understandingsociety.org.uk/publications/working-paper/2008-03.pdf
http://research.understandingsociety.org.uk/publications/working-paper/2010-04.pdf
http://research.understandingsociety.org.uk/publications/working-paper/2011-05.pdf
[3] http://www.esds.ac.uk/findingData/ukhlsTitles.asp

longitudinal survey data for academic and policy research across different disciplines. The use of geo-coded linked data enables greater research on 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 *Understanding Society* 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 *Understanding Society* 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 *Understanding Society* are available in Lynn (2009).[4] From Wave 2, the BHPS sample has been incorporated into the *Understanding Society* sample. The BHPS sample are interviewed in the first year of each wave.

3 Innovation Panel Wave 4: Design

IP4 was a face-to-face survey, with interviewers administering the survey using laptops (CAPI). The interviewing fieldwork ran from the 8th March to the 30th May 2011, followed by a re-issue phase running from 10th to 30th May. The initial issue period was extended because of lower-than-expected response. As part of an "early bird" experiment (see section 4c below), there was a two-week period before the CAPI interviewing fieldwork date (22nd February to 7th March) during which time sample members could contact interviewers to make appointments for interview from 8th March.

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

Prior to the survey going into the field there were eight 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, London (2 briefings) and Manchester. In total, 118 interviewers were briefed to work on IP4. Eight interviewers also attended a de-brief session in Brentwood on 2nd June 2011. The questionnaires used at IP4 are available from the *Understanding Society* website.[5]

a. Call for experiments

IP4 was the second time the Innovation Panel was open for researchers outside the scientific team of *Understanding Society* to propose experiments. A public call for proposals was made on 14th May 2010. Twenty-four proposals were received – more than double that which were received for IP3. Twelve were initially accepted although two were subsequently dropped. At the IP3 call a proposal had been received which had been held over for IP4. This resulted in eleven experiments being included at IP4. Submissions came from within ISER (fifteen), ISER in collaboration with other researchers (three) and from outside ISER completely (six). Of those that were external to ISER, three were from institutions within the UK, two from the US and one a collaboration between researchers from the US and the Netherlands. The twenty-four proposals were reviewed by a panel which included two ISER-based members of the *Understanding Society* scientific leadership team, and two members of the Methodology Advisory Committee to *Understanding Society* who were external to ISER.

b. Sample

There were two samples issued at IP4; the 'original' sample, and the refreshment sample. The original sample at IP4 comprised those household who had responded at IP3, plus some households which had not responded at IP3. Households which had adamantly refused or were deemed to be mentally or physically incapable of giving an interview were withdrawn from the sample. 1,354 original sample households were issued at IP4; 1,026 were previous-wave responding households, 102 were non-contacts at IP3, 157 were refusals at IP3 and 69 were non-responding for other reasons.

In addition to the original IP sample, a refreshment sample was issued at IP4. After three waves of experimentation, including a mixed-mode wave, the sample size of the Innovation Panel had decreased to just over 1,000 households. The refreshment sample was designed to bring the productive sample up to the original size of 1,500 households. The refreshment sample was a PAF sample (drawn from the small user Postcode Address File) of 960 addresses, 8 addresses in each of the 120 postal sectors from which the original sample was selected. In total, then, 2,314 addresses were issued to field at IP4.

[5] http://data.understandingsociety.org.uk/documentation/innovation-panel/questionnaires

c. Interview schedule

The questionnaire at IP4 followed the standard format used in the previous Innovation Panels as well as the main-stage of *Understanding Society*. The interview included:

- Household roster and household questionnaire: 15 minutes per household
- Individual questionnaire: on average 33 minutes for each person aged 16 or over
- Adult self-completion: around 9-14 minutes, paper questionnaire or computer self-administered interview (CASI)
- 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.

Some parts of the IP4 interview were recorded, using the lap-tops. This was done with permission of the respondent. The recordings were primarily of 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 68% of those who participated agreed to the sound recording. There was almost no difference in the levels of consent to recording between the original sample (68.7%) and the refreshment sample (67.6%).

d. Response rates

This section sets out the response rates for IP4 as a whole. Section 4f describes the effect of incentives on response rates. There were 2,314 households issued to field, 190 of those were found to be ineligible at IP4, whilst there were 41 ‘split-offs’, where one or more individuals had left the issued household and moved elsewhere, leaving sample members in the original household. This leaves an eligible sample of 2,165 households. Overall, just under two-thirds of households responded at IP4, that is, there was a household interview and at least one individual adult interview. However, this was the fourth time that the households had been approached for the original sample, but just the first time for the refreshment sample and so the overall response rate conceals differences between these two samples.

Table 1, below, shows the household response rates for the IP4 eligible sample. For the original sample, we also give the response rate for those households that responded at IP3 and those households that were not. A responding household is one in which the household questionnaire was completed and at least one eligible adult was interviewed.

Table 1: Household response at IP4

	Refreshment sample	IP4 original sample	IP3 responding	IP3 non-responding	Total
Responding	54.4%	69.9%	78.2%	34.1%	63.8%
	465	916	831	85	1,381
Non-contact	4.8%	5.4%	3.1%	15.3%	5.2%
	41	71	33	38	112
Refusals	39.1%	22.8%	17.0%	47.4%	29.2%
	334	299	181	118	633
Other non-responding	1.6%	1.9%	1.6%	3.2%	1.8%
	14	25	17	8	39
n	854	1,311	1,062	249	2,165

For the original sample, almost seven in ten households responded (69.9%). The level of refusals in the original sample was quite high, at 22.8%. This may be a consequence of issuing all non-adamant refusals from IP3, although the refusal rate for those households that had participated at IP3 was also relatively high (17%). As we would expect, those household that participated at IP3 were more likely to take part at IP4 than those which were non-responding at the previous wave. Over three-quarters of productive households at IP3 were productive at IP4 (78.2%). The policy of issuing previous-wave unproductive households, which incurs additional cost and increases the burden on interviewers, does result in a number of households being brought back into the responding sample. Just over one-third of households who were issued to field at IP4, despite being non-responding at IP3, responded. This demonstrates the value of issuing previous-wave non-responding cases on a longitudinal survey.

This was the first time the households in the refreshment sample were approached to take part in the IP. The response rate for this sample was slightly disappointing, the response rate at IP1 (in 2008) had been 59% and this was the target for the refreshment sample, which was located in the same PSUs as the original IP1 sample. However, the response after two re-issue periods was 54.4%.

Once a household responded, that is that the household questionnaire was completed and at least one eligible adult completed their individual interview, there does not appear to be much difference between the original IP sample and the new refreshment sample. Within responding households eight in ten eligible adults gave a full interview in both samples.

Table 2: Individual response within responding households at IP4

	Original sample	Refreshment sample	Total
Full interview	81.6%	81.0%	81.4%
	1,452	723	2,175
Proxy interview	8.5%	6.0%	7.7%
	152	54	206
Non-contact	1.6%	1.6%	1.6%
	28	14	42
Refusal	6.7%	8.6%	7.3%
	119	77	196
Other non-response	1.6%	2.8%	2.0%
	29	25	54
n	1,780	893	2,673

4 Experimentation in IP4

The experiments carried on IP4 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 previous waves of the IP. This section outlines the experiments carried at IP4; 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. The analyses in this working paper were based on a preliminary data-set which contained all cases but did not have weights or derived variables. The authors of each sub-section below are given in the heading.

a. Branched versus un-branched questions (Emily Gilbert)

Experiment proposers: Emily Gilbert, Nick Allum, Annette Jäckle

This experiment was designed to test for differences in measurement between branched and un-branched attitude scales. This is a repeat of an experiment carried at IP3, and cross-sectional results are available in the IP3 working paper. At IP4 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. The original sample were allocated to the same treatment as they had received at IP3. The refreshment sample were randomly allocated to each treatment. Two batteries of questions were used, concerning political efficacy and neighbourhood social cohesion. These are the same questions that were used at IP3. This experiment will be repeated once again at IP5 and longitudinal analyses using the IP3-IP5 data will be summarised in the IP5 working paper and be presented in more detail in subsequent publications by the authors of this experiment.

b. Re-issue Experiment (Sarah Budd)

Experiment proposers: Jonathan Burton

The re-issue period follows the main fieldwork period. It is a four-week period during which non-responding cases are sent back into the field for a second attempt at gaining co-operation. Re-issued cases include households where contact was not established during the initial fieldwork period and those that refused to participate (but not firm, adamant refusals). Ineligible households are not re-issued to field (non-residential addresses, vacant or derelict properties).

Re-issued households are sent a standard one page re-issue letter which simply informs them that an interviewer has been trying to get in contact[6]. This experiment was designed to see whether including additional information about the purpose and benefits of *Understanding Society* could improve response in the re-issue period.

Households were randomly assigned to one of two experimental treatments: A) standard reissue letter, or B) standard letter plus additional 'purpose' leaflet. The leaflet aimed to respond to common reasons for refusal and answer any questions the respondent may have about participating in the survey. 25 per cent of sampled households were re-issued at IP4 (581), 314 households received the standard re-issue letter and 267 received the additional re-issue materials. Seven percent of the re-issued households received a final outcome code of ineligible and have therefore been excluded from this analysis.

Table 3 shows re-issue outcome by experimental group. The proportion of households that were productive following re-issue was 31 per cent in group A compared with 27 per cent in group B. In other words, the households that received additional information in the re-issue letter appear to be *less* likely to be co-operative than the ones that received the standard letter. Households that received the additional re-issue material appear to be also slightly more likely to refuse participation (66% vs 62%). Non-contact during the re-issue period was about the same for both experimental groups at 7 per cent. However, none of the differences described above were statistically significant ($p > 0.05$).

Table 3: Re-issue outcome by experimental group

Re-issue group	Re-issue outcome		
	Productive	Refusal	Non-Contact
Group A (standard)	31%	62%	7.5%
Group B (additional)	27%	66%	7%
n	*158*	*345*	*39*

1. Base is eligible households that were reissued

This experiment has shown no significant difference in the conversion rate of households receiving a standard re-issue letter and households receiving additional 'survey purpose' materials.

[6] For copies of the fieldwork documents, see http://data.understandingsociety.org.uk/documentation/innovation-panel/fieldwork-documents

c. Want to be an Early Bird? Can encouraging respondents to contact interviewers to make appointments increase co-operation and save costs? (Matt Brown and Lisa Calderwood)

Experiment proposers: Matt Brown and Lisa Calderwood

The National Longitudinal Surveys, in the US, found that encouraging respondents to make contact with the fieldwork agency to book an appointment by paying a financial incentive led to significant reductions in the amount of interviewer effort required to achieve a complete interview (Kochanek et al, 2010) . The cost-savings achieved by reducing the number of interviewer calls required were more than sufficient to cover the additional costs of paying the (extra) incentive. IP4 included an experiment to evaluate whether this approach could be successful in a UK context and to explore whether sample members could be motivated to be an "early-bird" (i.e. to contact their interviewer at the outset of fieldwork to arrange an appointment) without a financial incentive.

Households which had responded in IP3 were randomly assigned to three groups:

Table 4: Design of the experiment and number of cases in each group

	Experimental Group	N	%
1	Early Bird Offer – Incentives	347	31.9
2	Early Bird Offer – No incentives	366	33.6
3	Control Group - No Early Bird Offer	375	34.5
	Total	1088	

Advance letters were posted to the two 'early bird' treatment groups approximately three weeks before fieldwork began. All adult sample members within the household received a letter. These letters informed sample members that the next wave of the study was about to begin and explained that their household was being provided with "the opportunity to request an "Early Bird Appointment" by contacting your interviewer on their mobile phone before anyone else to arrange your interview at a time that best suits you". The mailing also contained a leaflet entitled "Want to be an Early Bird?" which provided more information about the early bird offer and explained to those in the incentive group that if they contacted the interviewer to pre-book an appointment then everyone in the household who is interviewed would get an extra £5 to say thank-you. The leaflet for the non-incentive early bird group explained that contacting the interviewer to pre-book their appointment will 'make your interviewer's life much easier as they will not have to make repeated telephone calls or visits to your home in order to try and reach you'. The leaflet for both groups specified a two-week window prior to the start of fieldwork in which sample members could get in touch with their interviewer to book an appointment[7]. During this two-week window, interviewers were instructed not to attempt to contact sample members. Sample members were able to pre-book their interview for any day within the first four weeks of fieldwork. Once fieldwork began interviewers began contacting respondents as per normal.

[7] For copies of the fieldwork documents, see http://data.understandingsociety.org.uk/documentation/innovation-panel/fieldwork-documents

The control group also received an advance letter which explained that the next wave of the study was about to begin but did not encourage sample members to contact the interviewer and did not include an early bird leaflet.

The take-up rate was fairly low, eight per cent of households in the Early Bird groups took up the offer (i.e., contacted the interviewer, made an appointment and kept it) and the take-up rate was significantly higher in the incentive group than in the no incentive group (10% compared with 6%) ($p<0.001$).

Table 5: Take-up of the early bird offer by treatment group

	Take-up rate		
Group	**N**	**%**	**Base**
Early Bird Offer – Incentives	36	10.4	347
Early Bird Offer – No	21	5.7	366
Any Early Bird Offer	**57**	**8.0**	**713**

Completing interviewing in households which took up the Early Bird offer took significantly fewer interviewer visits ($p<0.001$) than in households which were either offered but did not take up Early Bird or were not offered Early Bird (1.4 compared with 3.5) (Table 6). Amongst those who took up Early Bird there was no difference in the average number of visits between the incentive and non-incentive treatment group or between those offered Early Bird and not taking it up and those not offered Early Bird.

Table 6: Number of interviewer visits required to complete interviewing in households in which a productive household interview was achieved (by whether Early Bird taken up).

	N	Mean	Std.
Early Bird offer taken up			
With incentive	36	1.37	.731
No incentive	21	1.45	.999
All taking up offer	***57***	***1.40***	***.830***
Early Bird offer not taken up			
Early bird offered - with incentive	231	3.49	2.290
Early bird offered – no incentive	265	3.41	2.039
Early bird not offered	274	3.44	2.040
All not taking up offer	***770***	***3.45***	***2.115***
All	**825**	**3.31**	**2.117**

However, the low take-up rates meant that overall, households offered Early Bird required only very slightly fewer visits than those not offered Early Bird (3.2 visits compared to 3.4) and the difference was not significant (Table 7). In order to achieve significant fieldwork efficiencies it will be necessary to boost the take-up rates of the Early Bird offer.

Table 7: Number of interviewer visits to all households (by experimental group).

	All households			
	N	**Max**	**Mean**	**Std. Dev**
Early Bird Offer – Incentives	347	14	3.10	2.290
Early Bird Offer – No incentives	366	21	3.21	2.152
Any Early Bird Offer	**713**	**21**	**3.16**	**2.219**
Control group – No Early Bird	375	13	3.37	2.224

d. Asking Respondents about Mode Preference (Olena Kaminska)

Experiment proposers: Olena Kaminska and Peter Lynn

This experiment is concerned with people's mode preferences, i.e. whether they prefer to be interviewed via face-to-face, postal paper and pencil questionnaire, web, telephone or any other mode. If people have such preferences, knowing them may help us tailor the offered mode of interview to that which is preferred. Two main questions are of interest. First, can we measure mode preferences reliably; and second, can expressed mode preferences help us to predict response in different modes?

The mode preference experiment is implemented in two waves of IP, wave 4 and wave 5. In both waves respondents are asked about mode preferences using five questions. Two general questions ask respondents to pick their most and least preferred modes among four modes (face-to-face, telephone, postal self-completion and web). The other three specific questions, using a scale from 0 to 10, ask respondents to rate their likelihood of responding in the future if contacted in each of three modes: telephone, postal and web.

Overall, face-to-face is rated as the most preferred mode (54.6%; with only 2.5% rating it as least preferred) among the four modes; and telephone mode is rated as the least preferred mode (with only 1.2% saying that telephone is their preferred mode, and 59% rating it as the least preferred mode). Note, the questions are answered as part of a face-to-face interview. Telephone received also the lowest rated likelihood of participation if a respondent is contacted in this mode the following year (likelihood rated as 2.62 on average on a 0 to 10 scale if a person is contacted on the telephone, in comparison to 5.68 if contacted via post or 4.88 if contacted on web).

As we were concerned about possible context effects, we randomized the order of the two types of questions (general and specific). In form A the three specific questions asking about likelihood of participation in each mode were asked before the general most and least preferred mode questions. In form B the order was reversed – the specific questions followed the general questions.

Interestingly, we find that general questions are prone to context effects with face-to-face mode being more preferred and web less preferred if specific questions are asked before the general questions (table 8). Unlike for general questions, no significant context effect on specific questions is observed (table 9). This suggests that asking likelihood to respond in each separate mode is less

prone to context effects and may be more suitable measure of mode preference. Nevertheless, the most important question is the strength of prediction of future participation in different modes, which will be investigated at wave 5 of the IP.

Table 8. Context effect on general questions of most and least preferred modes

	Most preferred mode		Least preferred mode	
	form A	form B	form A	form B
Face to face	66.27	55.42	2.65	2.97
Telephone	1.74	0.86	64.05	67.46
Post	12.98	17.64	9.67	10.91
Web	19.01	26.08	23.63	18.66
	Pearson chi2(3) = 32.9; Pr = 0.000		Pearson chi2(3) = 8.18; Pr = 0.043	

Table 9. Context effects on mean scores of likelihood to respond for specific questions

Form	Telephone		Postal		Web	
A	2.73	(0.1)	5.66	(0.11)	4.78	(0.12)
B	2.58	(0.1)	5.78	(0.11)	5.07	(0.12)
	t(2155)=1.05; p=0.15		t(2160)=-0.83; p=0.41		t(2160)=-1.71; p=0.96	

e. Interviewer Observations (Olena Kaminska)

Experiment proposers: Olena Kaminska and Peter Lynn

In face-to-face interview surveys it has become common to gather additional information from interviewer observations before contact is made with sample members. Such variables, as dwelling type, quality of the dwelling and the area may be collected for respondents and non-respondents and therefore may provide additional help to control for nonresponse bias. Two household characteristics that could be strong predictors, both of variables of substantive interest and non-response propensity, are car ownership and the presence of children. However, there are no standard interviewer observation questions relating to these characteristics. This experiment is concerned with developing such questions.

IP4 refreshment sample addresses were randomly assigned to one of two versions of interviewer observation questions. In version A the interviewer was asked to provide a judgment; in version B the interviewer was asked for factual information (text in panel below). Version A was asked as part of mainstage wave 1 UKHLS data collection. Using these questions in a nonresponse correction model we noticed that at least for the question on cars and vans ‘definitely having or not having a car’ predicts household response, and ‘likely/unlikely to have a car’ predicts nonresponse. This led us to think that interviewers may change their answers after an interview. Version B was developed as a potential solution to this. We expected that by asking directly for observable signs (version B) we eliminate the temptation for an interviewer to change their answers after an interview. We also

eliminate the possibility that interviewers may feel they are being tested, as households without any sign of a car can easily have cars. Note, each interviewer had a chance to receive both forms, and there was no control for previous interviewing experience in UKHLS, so some interviewers may have previously encountered Version A (at least 15 months earlier).

Version A

Based on your observation, is it likely that this address contains one or more children aged under 10 (including babies)?

- Definitely has a child / children aged under 10
- Likely
- Unlikely
- Definitely does not have a child / children aged under 10
- Cannot tell from observation

Based on your observation, is it likely that this address has a car or van?

- Definitely has a car / van
- Likely
- Unlikely
- Definitely does not have a car / van
- Cannot tell from observation

Version B

Standing outside, can you observe any signs of children under 10 (including babies) at this address?

- Yes
- No

Standing outside, can you observe any signs of a car or van belonging to this address?

- Yes, likely belonging to this address
- Yes, unsure whether belonging to this address
- No

An important result is that version B almost eliminates ‘don’t know’ responses to the interviewer observation questions. While the children question in version A has 48% and the car question has 34% of ‘don’t know’ responses, version B results in less than 1% of such responses (table 10). This is simply because while interviewers still don’t know about actual presence of kids or ownership of a car or van, they know whether signs of these can be observed when standing outside.

Table 10: Distributions of interviewer observations for two experimental groups

Children under 10				Cars and Vans			
Version A		Version B		Version A		Version B	
definitely present	5.5	yes	9.0	definitely present	37.0	yes, probably of this HH	48.1
likely present	6.4	no	91.0	likely present	19.7	yes, unsure if of this HH	10.6
unlikely present	26.0			unlikely present	6.4	no	10.8
definitely not present	13.7			definitely not present	2.8		
Cannot tell / DK	48.3	DK	0.0	Cannot tell / DK	34.1	DK	0.5
total	422		424		422		424

Next, we are interested in how interviewer observations predict household response outcome depending on the version. For this comparison, we take out the 'don't know' response, considering the large difference found before. Table 11 shows the distribution for responding and non-responding households for version A. While the chi-square test is not significant for the question about children, the pattern for the question about cars shows that responding households are more likely to receive categories 'definitely present' and 'definitely not present' than non-responding households, while category 'unlikely' is more prevalent among non-responding households. This is consistent with our earlier findings from wave 1 of *Understanding Society*.

Table 11: Relationship of response outcome with a version A of interviewer observations

	Children Under 10		Car or Van	
	nonresp HH	resp HH	nonresp HH	resp HH
definitely present	6.6	13.4	50.0	59.9
likely present	12.1	12.6	28.3	30.8
unlikely present	51.7	49.6	17.9	4.7
definitely not present	29.7	24.4	3.8	4.7
	$chi^2(3) = 2.92$ Pr = 0.40		$chi^2(3) = 13.3$ Pr = 0.004	

The version B questions have a more plausible relationship with response outcome (Table 12). For the question about car and van presence the relationship follows a logical pattern, although it is not significant (note, sample size is 422): households having their own car are more likely to respond, and those with no signs of a car have a lower chance to respond.

Table 12. Relationship of response outcome with a version B of interviewer observations

	Children Under 10			Car or Van	
	nonresp HH	resp HH		nonresp HH	resp HH
yes	7.73	10.00	yes, probably of this HH	44.27	51.74
no	92.27	90.00	yes, unsure if of this HH's	10.42	10.87
			no	45.31	37.39
	chi2(1) = 0.66 Pr = 0.42			chi2(2) = 2.8 Pr = 0.24	

Further inspection of these questions and their relationship with other interviewer observations reveals that when a household lives in a detached or semi-detached dwelling, interviewers guess that such a house is likely to have or definitely has a car in 91% of cases. Nevertheless, for the same household types interviewers observe signs of a car (either belonging to a household or not) for only 74% of cases. It seems that interviewers may use household type to make a guess about car ownership even if they don't see any obvious signs of a car when standing outside.

f. Incentives (Jonathan Burton)

Experiment proposers: Jonathan Burton, Peter Lynn, Heather Laurie, Noah Uhrig

At IP4 there were three incentive amounts for the 'original sample'; adults received either £5, £10 or £5 rising to £10 if every eligible adult in their household participated. The incentive allocation was at the household level, so everyone in the household received the same treatment. The voucher was included with the advance letter sent to each adult who was known to be in the household from the previous wave, including those who did not give an interview at IP3. The incentive, for these sample members, was therefore unconditional. The first stage of the analysis looks at the household and individual response to IP4 by the amount of the incentive. The second stage will then look at the change in incentive since IP3. Although there are only three incentive amounts, there are eight different incentive groups, reflecting the different treatments over the course of the IP. There are three groups who have had the same level of incentive over each wave, three groups who have seen a decrease, one group who have had an increase and one group where the level of the incentive decreased and then increased again. The analysis of this change will be simplified by just taking into account the change between IP3 and IP4.

Additionally, at IP4 there was a refreshment sample. This was a new sample in the existing PSUs, for whom this was their first wave of IP. For this sample we experimented with incentives that were much higher than the traditional incentive amount for a survey in the UK. Households in the refreshment sample were randomly allocated to receive £10, £20 or £30. One voucher was included with the advance letter – this was a PAF sample and so we did not have the names of people living at the addresses. The advance letter informed the sample member that if other adults were living in the household, they would also receive a voucher on completion of the survey. Thus, for the refreshment sample, the incentive was a combination of unconditional (for one person in the household) and conditional (for other adults).

Table 13, below, shows the household response rate for the original sample, by amount of incentive. The £10 incentive appears to have resulted in a higher rate of fully responding households, these are households where every eligible adult gave a full interview. Those households where adults received a £5 voucher appear to be more likely than the other incentive groups to be partially responding, that is where the household has been enumerated, the household questionnaire completed and at least one eligible adult giving a full interview whilst other eligible adults do not respond. However, neither the £10 or the £5-rising-to-£10 group are statistically significantly different to the £5 group for any of the outcomes.

Table 13: Household response rate by incentive group – original sample

	Original sample			
	£5	£10	£5-to-£10	Total
Fully responding	51.6%	55.4%	51.7%	52.9%
	316	242	109	667
Partially responding	21.9%	17.9%	17.5%	19.8%
	134	78	37	249
Non-contact	1.8%	1.6%	1.9%	1.7%
	11	7	4	22
Refusal	22.7%	23.3%	27.0%	23.6%
	139	102	57	298
Other non-responding	2.1%	1.8%	1.9%	2.0%
	13	8	4	25
n	613	437	211	1,261

*** p<0.001 ** p<0.01 * p<0.05

There appears to be a much larger difference among incentive groups for the refreshment sample (Table 14 below). This is the first time these households have been approached to participate in the Innovation Panel. Less than half of those households who received a £10 unconditional voucher responded (46.6%) whilst those who received a £30 voucher had a response rate that was 14.4 percentage points higher. The proportion of households who were partially responding in both the £10 and £30 groups are very similar, the main difference is in the proportion of fully responding households; 30.1% in the £10 group and 45.0% in the £30 group. This suggests that if it is important to get a response from every eligible person in the household, a larger incentive is likely to help. This is also suggested in the original sample, although the differences are not statistically significant, where those households who initially receive £5 for each adult are less likely to be fully responding that those who initially receive £10 each.

The increase in the household response rate is largely through the decrease in the proportion of whole-household refusals; from just under one-half in the £10 group to just over one-third in the £30 group. The £20 group falls between the £10 and £30 groups in the proportion of fully responding households, over 10 percentage points higher than the £10 group and just around 3 percentage points lower than the £30 group. The £20 group has the lowest rate of partially responding households by a small margin (not statistically significant). In terms of respondent households, either fully or partially responding, the £30 group is most successful (61.0%) followed

by the £20 group (56.5%) and with the £10 group the least successful (46.6%). The table below indicates the combination of incentive group and outcome which are statistically significantly different to the £10 group. There are no significant differences between the £20 and £30 groups. This suggests that to maximise response whilst keeping costs down, the £20 incentive may be more effective than the £30 incentive.

Table 14: Household response rate by incentive group – refreshment sample

	Refreshment sample			
	£10	£20	£30	Total
Fully responding	30.1%	41.7%**	45.0%***	39.0%
	84	118	129	331
Partially responding	16.5%	14.8%	16.0%	15.8%
	46	42	46	134
Non-contact	5.7%	3.9%	3.5%	4.4%
	16	11	10	37
Refusal	46.6%	37.5%	33.8%**	39.2%
	130	106	97	333
Other non-responding	1.1%	2.1%	1.7%	1.7%
	3	6	5	14
n	279	283	287	849

Tests of significance of the £20 and £30 incentive groups compared to the £10 incentive group.
*** p<0.001 ** p<0.01 * p<0.05

Table 15, below, shows the individual-level response within responding households in the original sample. Those individuals which received a £10 voucher were more likely to give a full interview and less likely to refuse than those who received £5.

Table 15: Individual response rate within responding households by incentive group – original sample

	Original sample			
	£5	£10	£5-to-£10	Total
Full interview	79.4%	83.5%*	84.0%	81.6%
	695	536	221	1,452
Proxy interview	9.1%	8.7%	6.1%	8.5%
	80	56	16	152
Non-contact	2.1%	1.6%	--	1.6%
	18	10	--	28
Refusal	7.9%	4.7%*	7.6%	6.7%
	69	30	20	119
Other non-responding	1.5%	1.6%	2.3%	1.6%
	13	10	6	29
n	875	642	263	1,780

Tests of significance of the £10 and £5-to-£10 incentive groups compared to the £5 incentive group.

*** p<0.001 ** p<0.01 * p<0.05

In the refreshment sample, once an interviewer has been successful in contacting and persuading someone in the household to participate, it appears that the level of the incentive plays much less of a role at the individual-level. There is no statistically significant difference between the incentive groups in the proportions of individuals, within responding households, who give a full interview. Individuals who received £20 are more likely to be non-contacts than the £30 group and less likely to be refusals than the £10 group. It should be noted, here, the relatively small sample sizes for these analyses.

Table 16: Individual response rate within responding households by incentive group – refreshment sample

	Refreshment sample			
	£10	£20	£30	Total
Full interview	77.5%	81.6%	82.9%	81.0%
	189	252	282	723
Proxy interview	5.7%	7.4%	5.0%	6.1%
	14	23	17	54
Non-contact[1]	0.8%	3.2%*	0.6%	1.6%
	2	10	2	14
Refusal[2]	12.3%	6.5%*	7.9%	8.6%
	30	20	27	77
Other non-responding	3.7%	1.3%	3.5%	2.8%
	9	4	12	25
n	244	309	340	893

1 Tests of significance of the £20 group compared to the £30 group

2 Tests of significance of the £20 and £30 incentive group compared to the £10 incentive group.

*** p<0.001 ** p<0.01 * p<0.05

For analysts of longitudinal studies it is important to have multiple observations through time for the same individuals. At a minimum two observations are required to enable researchers to analyse change over time. The re-interview rate at IP4, that is the proportion of those who gave a full interview at IP3 who also gave a full interview at IP4, is therefore a key indicator. Table 17, below, indicates that there are 1,205 adults who gave a full interview at both IP3 and IP4. This analysis only uses the original sample. There were no differences in re-interview rate by incentive group. However, individuals who were in the £5-rising-to-£10 group were more likely to be in a household which refused to participate at IP4 than those who were in either the £5 or the £10 group.

Table 17: IP4 outcome for adults who gave a full interview at IP3, by incentive group

	Original sample			
	£5	£10	£5-to-£10	Total
Full interview	76.5%	75.6%	73.0%	75.6%
	567	446	192	1,205
Proxy interview	2.4%	2.5%	0.8%	2.2%
	18	15	2	35
Non-contact	0.7%	0.3%	--	0.4%
	5	2	--	7
Refusal	1.5%	1.5%	0.4%	1.3%
	11	9	1	21
Other non-responding individual	0.9%	0.5%	0.8%	0.8%
	7	3	2	12
Household non-contact	0.5%	0.3%	0.8%	0.5%
	4	2	2	8
Household refusal	12.8%	14.9%*[1]	21.7%**[2]	15.1%
	95	88	57	240
Household other non-responding	0.8%	1.0%	0.8%	0.9%
	6	6	2	14
Untraced mover	3.1%	1.7%	1.5%	2.3%
	23	10	4	37
Household no longer eligible	0.7%	1.5%	0.4%	0.9%
	5	9	1	15
n	741	590	263	1,594

1 Tests of significance of the £10 incentive group compared to the £5-to-£10 incentive group.
2 Tests of significance of the £5-to-£10 incentive group compared to the £5 incentive group.
*** p<0.001 ** p<0.01 * p<0.05

In addition to analysing the amount of the incentive, we can also look at the effect of a change in the level of the incentive. At IP4 two of the eight incentive groups in the original sample had their incentive increased from £5 to £10 whilst the other six groups had the same incentive amount as at IP3. The proportion of households in the groups which saw an increase that were fully responding was 53.8%, compared to 52.6% for those with no change in incentive level. This is not a statistically significant difference. When those households which are partially responding are included, the household response rate was 71.6% for the households with an increased incentive compared to 73% for those with an unchanged level of incentive. At the individual level, the re-interview rate for those who had seen an increase in the incentive (73.5%) was not statistically different to those who had not (76.4%).

Of the two groups that had an increase in the level of incentive, one had already seen a decrease from £10 at IP1 to £5 at IP2 and IP3, before the increase back to £10 at IP4. For the other group, this was the first time it had been £10, having been at £5 for all the previous IP waves. For this latter group, the household response rate was 54.9%, compared to 51.4% for the group who had previously seen the decrease, again this is not a statistically significant difference.

g. Consent to data linkage (Jonathan Burton)

Experiment proposers: Jonathan Burton, Gundi Knies and Emanuela Sala

At the first wave of the Innovation Panel (IP1), in 2008, a sub-sample of respondents were asked for their permission to link to administrative information held by the Department for Education (DfE) and the Department for Work and Pensions (DWP). Those who had not given consent (or were not asked) for linkage to education records at IP1 were asked to give consent at IP2. In addition, at IP2, consent was sought from all respondents for linkage to health administrative records.

Ethical guidance in the UK requires us to periodically give sample members a chance to change their minds about consent. Sample members are able to revoke their consent at any time by contacting ISER in writing, either using a Freepost address or email. In addition to this, however, we must actively give sample members an opportunity to change their mind. There are a number of ways this can be done and different data holders may have different preferences. For example, on the UHKLS between waves 3 and 4, those who gave consent to link administrative health records to their survey responses received a letter reminding them of this consent and enabling them to return a slip in a freepost envelope if they wished to revoke their consent. During Wave 4, those who gave consent to education data linkage were asked to confirm their consent as part of the interview, with those who choose not to confirm consent being treated as if they had revoked it. In the case of both health and education, those who had not given consent in the past and new entrants to the study were asked a standard request for consent during the interview.

At IP4 we wanted to analyse more systematically whether (a) the format and (b) the placement of a question to (re-)ask for consent affected the consent rate, that is the proportion of the sample who give consent. IP4, therefore, included a request to all respondents for consent to link administrative information held by the DWP to the survey answers they gave. We found that both the placement of the request, and the way in which the request was phrased, had a significant effect on consent rates.

To test different ways of phrasing the consent question(s) during the interview the sample was randomly allocated to two groups. One group were reminded of their previous answer (at IP1) and asked whether they would consent to data linkage ("dependent" question). The other group were not reminded of their previous response but were asked the question 'afresh', in the same way as those who had never been asked before ("independent" question).

Our hypothesis was that those who were reminded of their previous response would be more likely to give the same response as before, compared to those who were being asked independently. There may be at least two reasons why we might expect this behaviour. Firstly, there is evidence that people like to be consistent when they are responding to survey questions (Groves, Cialdini and Couper, 1992). If they answer in a contradictory way, they may appear to the interviewer to be indecisive or give the impression that they had answered 'wrongly' at the previous opportunity. Thus, by reminding respondents of their IP1 response, they are likely to give the same answer at IP4. When respondents are asked the question independently, they have no reminder of their previous answer and so they are able to make the decision at the time without worrying about consistency. Secondly, the decision whether or not to consent to data linkage may be quite difficult.

Whilst there are some who are likely to have a 'gut reaction' and so not require much thought, there will be others for whom the answering the question may require cognitive effort. The information that the interviewer gives the respondent, including a leaflet about data linkage, gives the respondent additional information to process. By reminding the respondent of their previous response we give them an opportunity to short-cut this cognitive process by giving them an easy response; to agree with their previous answer. This short-cut is justified by the respondent 'trusting' their earlier thought processes, rather than thinking through their response from the beginning. Thus, we are more likely to see agreement with IP1 responses for those who are reminded of their previous response than when the question is asked independently.

The table below shows the text of the question for these treatment groups.

Table 18: Consent question wording for experimental treatment groups. Dependent and independent question by previous wave consent

<table>
<tr><th></th><th>Independent question</th><th>Dependent question</th></tr>
<tr><td>Gave consent at IP1</td><td rowspan="3">Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. Are you happy for us to do so?</td><td>Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. According to our records, when we interviewed you in 2008, you gave us permission to do so. Are you still happy for us to do so?</td></tr>
<tr><td>Did not give consent at IP1</td><td>Finally, we would like to add information on your National Insurance contributions, benefits and taxes, savings and pensions from administrative records held by the DWP to your survey responses. According to our records, when we interviewed you in 2008, you declined that we do this. Are you willing to give your consent now?</td></tr>
<tr><td>Not asked for consent at IP1</td><td>Not applicable</td></tr>
</table>

The sample sizes for the analysis of the question wording and the achieved consent rates (the proportion of those asked who gave consent to data linkage) for each group are shown in the table below. It should be noted that the IP4 refreshment sample and those who had been new entrants to the original sample since IP1 were all asked for consent using the independent question, since they had not been part of the study at IP1. These sample members are not included in the analysis to evaluate the effect of dependent or independent question wording but are included in the analysis of the placement of the consent question.

Around six in ten (61.6%) of those who were independently asked for consent to link to benefit data gave their consent. This proportion was similar across the samples that were asked independently;

60.1% for those in the original sample who had been asked for consent at IP1, 65.6% for those in the original sample who had not been asked at IP1 or were new entrants to the sample since IP1 and 60.7% for those in the refreshment sample, for whom this was the first time they had been part of the IP.

For those who were reminded that they had given consent at IP1, the consent rate was almost nine in ten (89.1%). Only slightly more than two in ten respondents who were reminded that they had declined to give consent at IP1 gave consent at IP4 (22.4%). We know whether those who were asked the independent question had given consent at IP1, so we can compare those who agreed (refused) at IP1 and were asked the independent question at IP4 with the percentage of those who agreed (refused) in the dependent question.

Those who gave consent at IP1 and were reminded of this at IP4 were almost 20 percentage points more likely to give consent at IP4 than those who were not reminded (89.1% compared to 69.5%). Similarly, those who refused to give consent at IP1 were almost half as likely to consent at IP4 when reminded about their previous response (22.4% compared to 44.4%). Even with the relatively small sample sizes for these groups, the differences are statistically significant.

Table 19: Sample sizes and consent rates for experimental treatment groups. Dependent and independent question by previous wave consent

	Independent question		Dependent question		
	Sample size	Consent rate	Sample size	Consent rate	n
Gave consent at IP1	315	69.5	359	89.1***	674
Did not give consent at IP1	189	44.4	219	22.4***	408
Not asked for consent at IP1	360	65.6	4[1]	50.0	364
Refreshment sample	722	60.7	n/a	n/a	722
Overall	1,586	61.6	582	63.8	2,168

*** p<0.001 statistical significance test of dependent question versus independent question

[1] There were 4 people in the original sample that were not asked the consent question at IP1 but were routed through the dependent question in error.

The second experimental treatment was in the placement of the consent question in the interview. Our hypothesis was that if the request for consent appeared in a context which made the request more salient, the respondent would be more likely to agree to data linkage. We were asking for permission to link to administrative data held by the DWP, primarily information about the receipt of state benefits. Around half of respondents were asked for consent after a module of questions which asked about the receipt of state benefits and other payments. The other half of respondents were asked for their consent at the end of the questionnaire. Those respondents asked at the end of the questionnaire would have been through the benefits module, and then asked about savings, personal pensions, retirement planning, politics, political self-efficacy, political networks, environmental behaviours and attitudes and a self-completion (CASI) section before reaching the consent question.

The question which appeared after the benefits module started with the phrase "We would also like to add information on your National Insurance contributions…", whilst the question at the end of the questionnaire is shown in table 18 above, starting with "Finally, we would like to add information on your National Insurance contributions…". Other than this, the questions were identical. The allocation to this experiment was done independently to the independent/ dependent interviewing question experiment. The allocation was also independent of whether the respondent was in the original sample or the refreshment sample. The refreshment sample are therefore included in the table below. Those who were asked for their consent after the benefits module were almost 5 percentage points more likely to give consent than those asked at the end of the questionnaire.

Table 20: Consent rates by placement of the request

	Gave consent
Question in context	64.7%
Question at the end	59.8%*
n	2,168

* $p<0.05$ statistical significance text of question at the end compared to question in context

Another explanation for this finding is that towards the end of the questionnaire the respondent may want to hurry the interview along because of the length of the questionnaire. They may therefore be less inclined to spend time reading an information leaflet and consent form and giving the matter their full consideration. Unfortunately, we cannot disentangle this explanation from the contextual explanation, since the benefits module always appeared at the same place in each interview. To test this, it would be necessary to manipulate the placement of the benefits module.

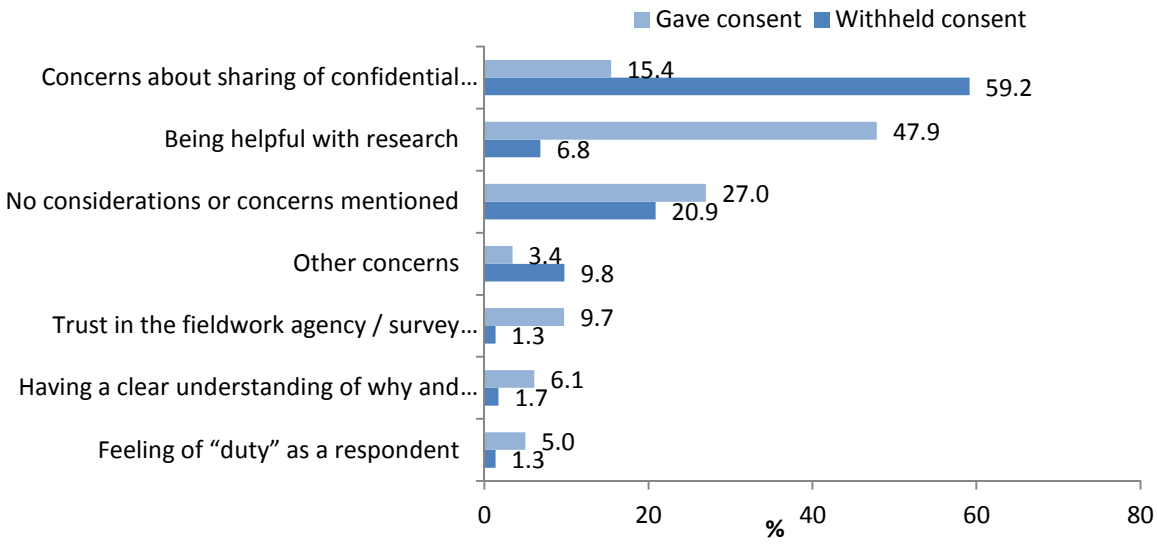
As part of IP4 we also asked the interviewer whether they thought that anyone in the household had influenced the respondent's decision at the consent question. Around one-in-twenty (5.3%) of respondents were judged to have been influenced. Those who were influenced were less likely to give consent; 50.9% gave consent compared to 62.8% of those who were not influenced ($p<0.05$).

Regardless of whether the respondent was asked the consent question in context or at the end of the questionnaire, the interviewer was required to get the respondent to read and sign a consent form. This process happened at the end of the questionnaire, so as not to disrupt the flow of the interview. At this point the interviewer coded whether or not the respondent had changed their mind. Altogether 4.2% of respondents changed their mind, however those who had originally given consent were more likely to then refuse at the time of signing the consent form (5.4%) than those who had initially decided against giving consent changing their mind and signing the form (2.3%). There was no difference in the proportions who changed their mind by whether the question had appeared earlier in the interview or later. This suggests that asking earlier in the interview, and then requiring a signed form at the end of the interview doesn't reduce consent rates through people changing their mind during the rest of the interview.

Regardless of whether the respondents gave or withheld their consent, they were asked about what they were considering when they gave their response. We were interested in the motives that respondents would give to explain why they did, or did not, consent to data linkage. The question asked was "Different things can be important when deciding to give consent to add information from DWP administrative records to survey data. What were you considering when answering?".

Just under a third of respondents mentioned that they considered being 'helpful' with research (32.2%), a similar proportion said they had concerns about sharing confidential information (31.9%). Not surprisingly, wanting to be helpful with research was cited more often by those who consented (47.9%) than those who declined to consent (6.8%). Conversely, concern about confidentiality was more often cited by those who did not consent (59.2%) than those who did (15.4%). It is worth noting that just over one in six people who consented still had concerns about confidentiality, and this demonstrates a need to raise the issue but to try and reassure respondents that their information will be kept confidential. It suggests that if interviewers are able to reassure respondents, concerns about confidentiality need not lead to a refusal to consent. Just under one in ten of those who gave consent said it was because they trusted the fieldwork agency or survey organisation. A similar proportion (9.8%) of those who withheld consent had 'other concerns', compared to 3.4% of those who consented. Among those who gave consent 6.1% said it was because they clearly understood why and how the linkage would take place, compared to just 1.7% of those who declined to consent. This demonstrates the importance that the reasons why linking survey responses to administrative data helps research, and the process by which the information is linked, are available to the respondent. Just one in twenty of those who gave consent said that they considered it their 'duty' as a respondent (5%). Just over a fifth of those who declined to give consent (20.9%) and just over a quarter of those who gave consent (27.0%) did not give any concerns or considerations about the decision. The graph below shows the proportions of those who gave consent and those who withheld consent who gave a response. Respondents could name more than one reason, and the interviewer coded their response to a pre-set list of categories with two "other" categories where the reason was recorded by the interviewer verbatim. The categories on the graph below are those where the difference between the two groups are statistically significant ($p<0.01$).

Graph 1: Rationales for giving or withholding consent, differentiated by consent outcome.



After asking for the concerns and considerations the respondent had when considering the request to link to administrative data the interviewer coded whether the respondent had then changed their mind about their decision over whether to give consent. Just over 4% of respondents did change their mind, with similar proportions among those who had been asked earlier in the interview (4.6%) and those who were asked just a minute or so before (3.9%). Those who had initially indicated their consent were more likely to change their mind (5.4%) than those who initially said they did not consent (2.3%).

We find, then, that the placement of the consent question has an effect on whether people give consent. If the question is placed in a context where the request is salient, people are more likely to give consent. In this experiment we manipulated the placement of the request to link to benefit data to come after a module of questions on benefit receipt. Future research could test to see whether this finding holds for other domains of administrative data, for example requesting linkage to health administrative data after a module of questions on health and use of health services, or linkage to education administrative data after questions about school-level, further and higher education qualifications. When having to re-collect consent to data linkage, we have shown that reminding people of their earlier decisions prompts them to make the same decision. Thus, to maximise the number of people for whom consent is retained, a strategy would be to remind those who had previously given consent whilst those who have not given consent in the past are asked an independent question. Finally, we find that there are a number of reasons that respondents give for consenting or not, but that those who have concerns may still be reassured and willingly give their consent. Concerns about confidentiality, though, is the main reason given by those who withhold consent and so improving messages about data security may be important in easing these concerns.

h. IP4 Context effects (Noah Uhrig)

Experiment proposers: Noah Uhrig, Annette Jäckle, Emanuela Sala, Fred Conrad

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 21 shows the response distributions for the target questions at IP3, 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' (6.3%), than respondents in the high frequency context group (3.0%) and less likely to say never (9.1%) versus (13.8%). This general pattern is repeated across the remaining target questions, though does not reach conventional levels of statistical significance apart from "fellow workers". There, the low frequency context respondents were significantly more likely to say 'always' (4.1% vs 0.9%) and less likely to say 'never (19.0% vs 23.8%).

Inspection of the same analysis of data from wave 4 of the Innovation Panel finds no effects of context across any of the target items. That is, the expected pattern of low frequency context being associated with a greater likelihood of saying 'always' and less likelihood of saying 'never' was not consistently displayed across target items. Table 22 shows no significant associations between context and responses. Nevertheless, the hypothesized longitudinal effect of change in context is that respondents who switch context between waves are likely to be affected by context more than those who do not switch contexts. Table 23 contains results of this analysis.

Table 23 shows generally across target items that respondents experiencing the same context are more likely to increase frequency of discussing politics than those who switch context. And, those

who switch context are more likely to decrease frequency of discussing politics than those who experienced the same context. Only the association between context and change in response for the target item 'other family members' achieves any level of conventional statistical significance, however. There, respondents experiencing the same context were more likely to either increase frequency or answer the same as at the previous wave (29.5% and 43.6% vs 25.6% and 40.2%) while those who swapped contexts were more likely to decrease frequency (34.2% vs 26.9%). Upon a more detailed inspection, change in response seemed to depend largely on the initial context. There was no significant change in the types of answers given by respondents at IP4 who had experienced the high frequency context at IP3. However, respondents who experienced the low frequency context at IP3 were more likely to be effected by change in context at IP4. Those who experienced the same low frequency context at IP3 and IP4 were more likely to increase the frequency of talking about politics with family members (33.2% vs 24.7%) whilst those who swapped context from low to high frequency were more likely to decrease frequency of talking about politics with family members (39.2% vs 22.3%).

These initial analyses suggest that context does matter and that switching contexts between waves of a longitudinal study could influence the response process over the life of a panel. However these data require more robust analysis as there are unanswered questions.

Table 21: Context effects in IP3 (row %)

Talks about politics to…	Context	Always	Very often	Quite often	Sometimes	Rarely	Never	N (weighted)	P (Chi^2)
Husband/wife/partner	Low freq	6.3	10.7	17.7	33.5	22.1	9.1	350.8	
	High freq	3.0	13.7	22.9	31.1	15.4	13.8	363.9	0.056
Family members	Low freq	3.0	7.1	10.6	27.5	28.5	23.2	561.2	
	High freq	2.2	7.3	12.8	28.4	24.9	24.5	603.5	0.770
Friends	Low freq	1.9	5.0	14.3	31.0	26.0	21.9	561.2	
	High freq	0.6	7.0	14.2	28.1	26.5	23.7	603.5	0.447
Fellow workers	Low freq	4.1	8.4	10.4	30.5	27.7	19.0	330.8	
	High freq	0.9	8.1	16.5	28.8	21.9	23.8	361.5	0.031
Neighbours	Low freq	0.8	0.9	2.6	7.5	24.6	63.6	561.2	
	High freq	0.3	0.3	1.9	8.1	24.6	64.8	603.5	0.651
Anyone else	Low freq	0.3	0.9	1.1	10.5	20.9	66.3	561.2	
	High freq	0.2	0.7	1.8	10.5	21.1	65.6	602.7	0.960

Notes: Analysis uses cross-sectional individual weights, all statistics adjusted for sample clustering.

Table 22 Context effects in IP4 (row %)

Talks about politics to…	Context	Always	Very often	Quite often	Sometimes	Rarely	Never	N (weighted)	P (Chi^2)
Husband/wife/partner	Low freq	3.8	11.9	21.1	26.8	21.4	15.0	371.1	
	High freq	3.2	14.2	18.9	29.4	24.8	09.6	370.5	0.324
Family members	Low freq	1.3	7.0	12.9	26.4	28.1	24.4	597.4	
	High freq	1.2	5.8	18.1	25.9	29.2	19.7	620.4	0.242
Friends	Low freq	1.6	7.3	12.9	28.1	27.8	22.4	597.4	
	High freq	0.5	7.1	18.9	27.6	26.6	19.3	620.4	0.144
Fellow workers	Low freq	2.8	8.9	16.7	23.4	22.1	26.1	340.1	
	High freq	0.6	8.4	13.4	29.9	25.7	22.1	353.1	0.160
Neighbours	Low freq	0.4	1.8	1.6	7.7	29.2	59.3	597.4	
	High freq	0.7	0.8	3.3	11.5	25.8	58.0	620.4	0.139
Anyone else	Low freq	0.4	1.0	2.6	10.9	21.2	63.8	596.7	
	High freq	0.3	1.5	1.6	12.7	24.3	59.6	620.4	0.547

Notes: Analysis uses cross-sectional individual weights, all statistics adjusted for sample clustering.

Table 23: Longitudinal Context Effects IP3 to IP4 (Row %)

Talks about politics to…	Context	Increasing frequency	No change	Decreasing frequency	N (weighted)	P (Chi^2)
Husband/wife/partner	Swapped context	55.7	24.3	20.1	527.9	
	Same context	58.0	26.0	16.1	487.9	0.355
Family members	Swapped context	25.6	40.2	34.2	527.9	
	Same context	29.5	43.6	26.9	487.9	0.089
Friends	Swapped context	22.7	44.3	33.0	527.9	
	Same context	27.7	39.6	32.7	487.9	0.258
Fellow workers	Swapped context	60.4	21.4	18.2	527.9	
	Same context	56.1	25.1	18.8	487.9	0.460
Neighbours	Swapped context	13.5	61.5	25.0	527.9	
	Same context	16.4	59.3	24.3	487.9	0.626
Anyone else	Swapped context	16.8	55.2	28.0	527.9	
	Same context	19.4	55.6	24.9	487.9	0.5847

Notes: Analysis uses longitudinal individual weights, all statistics adjusted for sample clustering.

i. Panel Conditioning in IP4 (Noah Uhrig)

Exerpiment proposers: Noah Uhrig

"Panel Conditioning"[8] is a term which suggests a systematic effect of panel participation on either response behaviour or on the behaviour targeted by survey measurement (Waterton and Lievesley 1989). If conditioning effects are large, then the utility of panel data is compromised (Cantor 1989; Cantor 2008). A number of studies have demonstrated that survey participation induces such diverse activities as voting, immunization and diagnosis of arthritis (Bartels 1999; Battaglia, Zell, and Ching 1996; Clausen 1968; Kraut and McConahay 1973; Wilson and Howell 2005; Yalch 1976). Other work suggests that panel participation affects opinion formation. For example, Sturgis and colleagues (2009) observe reduced variances in opinion items over several waves of a panel, an increase in wave-lagged correlations over time, and a decrease in "Don't Know" responses. On-going participation has also been shown to enhance accuracy in survey reports in other ways (Ferber 1953; Sturgis, Allum, and Brunton-Smith 2009; Waterton and Lievesley 1989). An alternative strand of work suggests that conditioning could inhibit accuracy in reporting (Bailar 1975; Cohen and Burt 1985; Ghangurde 1982; Kemsley 1961; Mooney 1962; Neter and Waksberg 1964; Silberstein and Jacobs 1989; Turner 1961). The extent to which on-going participation in a panel survey affects panel responses and subsequent respondent behaviour, therefore, remains an open question.

An experiment carried since IP1 of *Understanding Society* focuses on exposure to survey content as a potential mechanism leading to conditioning effects. At IP1 a random half of the sample was asked to self-report their height and weight, and to provide their opinion about the existence and likely effects of climate change. At IP2, the entire sample was asked to self-report height and weight, and to report on the frequency of engaging in various environmentally beneficial behaviours[9]. At IP3, the same random half of respondents as at IP1 where asked these questions while at IP4 the full sample again received this content. This experimental design means that half of the sample self-reported height, weight and environmental behaviour annually whereas the other half of the sample self-report on these things at the lower frequency of every other year. It is hypothesised that the greater the frequency of repeating content, the greater opportunity for conditioning effects to occur. If exposure to survey content enhances reporting accuracy, then known biases in self-reports of height and weight should be attenuated due to conditioning. And, if prior survey content induces a crystallisation in opinions and behaviour, then conditioned respondents should exhibit less variant environmental behaviour as opposed to unconditioned respondents.

[8] Other terms used to name the phenomenon include, but are not limited to, 'Time in Sample Bias', 'Rotation Group Bias', 'Repeated Interviewing Effect', 'Panel Membership Effect', and 'Panel Bias'. These terms often provide a clue as to how the author designs their analysis and may also indicate particular assumed causal mechanisms.

[9] Items include: Leaving TV on stand-by for the night; Switch lights off in rooms not used; Keep tap running while brush teeth; Add clothes rather than turn-up heat when cold; Not buy product with too much packaging; Buy recycled paper products; Use own shopping bag; Use public transport rather than car; Walk/cycle short journeys; Car share; Take fewer flights were possible. With response options: Always, Very Often; Quite Often; Not Very Often; and Never. A scale combining these items is standardised.

Results: Height and Weight. Analysis of IP1 and IP2 data suggest a conditioning effect for women, but not for men (see Uhrig 2012). Uhrig finds that conditioned women tend to report being heavier and taller than unconditioned women. Consequently, heavier conditioned women tend to have a calculated body-mass that is about one point higher than unconditioned women. In terms of data quality, conditioned women are less likely to provide rounded bodyweights, that is the whole or half stone, they are less likely to non-respond when asked for their body weight, and they are more likely to have weighed themselves recently. Uhrig's results for men are less consistent and largely insignificant. Taken together, analyses of IP1 and IP2 data suggest that conditioning enhances the reporting of accurate data for women only.

Comparing results from IP1 and IP2 with a repeated design at IP3 and IP4, we find all effects to be attenuated. The table below shows that at IP2, the 75th percentile for conditioned women was about a half-stone heavier than unconditioned women, but no effect for men. By IP4, there is no comparable statistically significant effect for women, though the results are of the same sign and magnitude.

Table 24: Results from simultaneous quantile-regressions of conditioning on the 25th, 50th and 75th percentiles of self-reported bodyweight. Shown are the effects of conditioning on the 75th percentile only.

	Men's 75th Percentile	Women's 75th Percentile
Conditioning effect at IP2	-4.48	7.04**
	(3.42)	(3.17)
Conditioning effect at IP4	-0.63	5.28
	(4.35)	(4.53)

** $p < 0.05$

Notes: Shown are coefficients. Standard errors, shown in parentheses, are obtained by bootstrap methods (500 replications) and are adjusted for clustering in sample design. Respondent age, education, reporting of rounded values and recent weighing are controlled in the models but are not shown.

Considering height, effects of conditioning on the 75th percentile for men are comparable between IP2 and IP4. Notably, the coefficients for conditioning are significant at both IP2 and IP4, implying that conditioned men are more likely to report being taller than unconditioned men. There are no effects of conditioning at either IP2 or IP4 for women's self-report of height.

Table 25: Results from simultaneous quantile-regressions of conditioning on th e25th, 50th and 75th percentiles of self-reported bodyweight. Shown are the effects of conditioning on the 25th and 75th percentiles of height and body-mass index for men and women.

	Men's 25th Percentile	Men's 75th Percentile	Women's 25th Percentile	Women's 75th Percentile
Conditioning effect on Height at IP2	0.16	0.62***	0.16	0.10
	(0.27)	(0.24)	(0.23)	(0.23)
Conditioning effect on Height at IP4	0.00	0.75**	0.21	0.39
	(0.36)	(0.37)	(0.27)	(0.36)
Conditioning effect on BMI at IP2	-0.41	-0.42	-0.11	0.93
	(0.40)	(0.43)	(0.39)	(0.66)
Conditioning effect on BMI at IP4	-0.71	0.21	-0.87**	-0.15
	(0.49)	(0.65)	(0.44)	(0.68)

** $p < 0.05$, *** $p < 0.01$

Notes: Shown are coefficients for conditioning only. Standard errors, shown in parentheses, are obtained by bootstrapped methods (500 replications) and are adjusted for clustering in sample design. Respondent age and education are included in the model, but not shown in the table.

There are no statistically significant effects of conditioning on men's body-mass index (BMI) at either IP2 or IP4. For women, there are also generally no statistically significant effects, however the 25th percentile for conditioned women is significantly lower than for unconditioned women ($\beta = -0.87$, $p < 0.05$).

Table 26: Results from logistic regression of conditioning and recent weighing on various indicators of survey response quality.

	Men			Women		
	Rounding	*Weight Non-Response*	*Recent Weighing*	*Rounding*	*Weight Non-Response*	*Recent Weighing*
Conditioning effects at IP2	0.02	0.05	-0.19	-0.08	-0.78	0.21
	(0.20)	(0.61)	(0.16)	(0.17)	(0.60)	(0.14)
Conditioning effects at IP4	-0.20	1.88*	-0.04	-0.02	-0.89*	-0.12
	(0.23)	(1.08)	(0.19)	(0.21)	(0.46)	(0.17)

* $p < 0.10$

Notes: Shown are coefficients. Age and education are controlled in the models but are not shown in the table. Standard errors, which are adjusted for clustering in sample design, are shown in parentheses.

Survey response quality could be enhanced by a reduction in the likelihood of estimating answers, or providing "Don't Know" responses. The table above reports the effects of conditioning at IP2 and IP4 on whether the respondent provided a rounded number when self-reporting body weight, answers "Don't Know" to the body weight question, or indicated that

they weighed themselves within the week prior to interview. We see that there is generally no effect of conditioning on any of these indicators of data quality at IP2, but by IP4 conditioned men were marginally more likely to answer "Don't Know" to the weight question than unconditioned men ($\beta = 1.88$, $p < 0.10$) and that conditioned women were marginally less likely to answer "Don't Know" to the weight question than unconditioned women ($\beta = -0.89$, $p < 0.10$).

Results: Environmental Behaviour. We formed a single standardised multi-item environmental behaviour scale at IP2 and again at IP4. At both time points, unconditioned respondents seem to have a *higher* average inter-item correlation on the scale (IP2: $\mu_{unconditioned} = 0.11$, $\mu_{conditioned} = 0.08$, $p < .0001$; IP4: $\mu_{unconditioned} = 0.11$, $\mu_{conditioned} = 0.09$, $p < .0001$). This means that the unconditioned sample tends to be more internally consistent in their answers across the items than the conditioned sample. A comparison of scale variances for each group shows slightly different results at IP2 and IP4. Notably, in IP2, the conditioned sample has a significantly lower variance ($VAR_{unconditioned} = 0.204$, $VAR_{conditioned} = 0.173$, $F = 1.18$, $df = 828, 812$, $p < 0.01$) whereas at IP4 there is no difference in variances across experimental treatments. This means that at IP2 unconditioned respondents were more different from one another in beliefs and values but by IP4 both groups were more comparable ($VAR_{unconditioned} = 0.199$, $VAR_{conditioned} = 0.192$). This latter finding could be due to two things. First, as the "unconditioned" sample get asked the questions a second time, they could revise their answers to be more socially consistent and possibly desirable – a process occurring earlier for conditioned respondents. Alternatively, selective attrition could account for increasing similarity between groups. In other words, overtime, respondents most interested in environmental issues tend to stay in the study insofar as these questions are highly salient and relevant for them. Respondents in both groups who are uninterested in environmental matters will attrite from the study. The next result is little or no difference across groups. This explanation implies differential rates of attrition by experimental treatment which is a one further area ripe for further investigation.

j. Between-wave mailing experiment (Jonathan Burton)

Experiment proposers: Jonathan Burton, Chris Garrington, Paul Groves

One of the challenges in a longitudinal study, such as *Understanding Society*, is keeping track of people who change addresses between waves of data collection. Sample members are sent change-of-address cards to return to ISER when they move house. By sending a card back with a new address, the sample member receives a £5 voucher as a token of appreciation. Incentivising sample members to inform ISER of their new address is more cost effective than having an interviewer carry out tracing activities in the field.

Understanding Society also has a Participants website. This website gives participants more information about the study, including news of recent findings and press coverage. As part of the between-wave mailings, sample members are sent unique log-in identifiers to allow them to register with the website. Those who register are able to inform ISER of a change of

address on the website, and also receive additional email newsletters. When registering, the sample member has to give an email address which is used to send a verification email with a link which needs to be clicked to activate the registration. This is an additional source of email addresses, which are useful as part of tracing procedures.

It is therefore in the interests of those managing the study to get as many sample members registering with the website as possible. At IP4 an experiment was proposed to measure the effectiveness of different methods of asking sample members to register with the website. There were four treatment groups; receiving the request in a letter or receiving it in an email, and being promised an incentive worth £5 or not being offered an incentive. These treatment groups were independent of each other and households were randomly allocated to each treatment. All adults in a household received the same treatment. There are therefore four groups; (i) letter and no incentive, (ii) email and no incentive, (iii) letter with offer of an incentive and (iv) email with offer of an incentive. The incentive was to be selected once the individual had registered and was a choice of a £5 gift voucher, £5 Amazon e-voucher[10], £5 iTunes e-voucher[11] or a £5 donation to either Help for Heroes[12], NSPCC[13] or Oxfam[14]. Once the initial allocation to treatment group was completed, those in the email groups for whom we did not have an email address were re-allocated to the equivalent non-email group.

The outcome of interest in this experiment is whether the sample member registered with the Participant website. In total, of 2631 adults who were mailed just 102 (3.9%) registered with the Participant website. Table 27 below shows the proportion of people who registered in each group. The upper part of the table uses the original allocations, before those in the email group without an email address were re-allocated. For the two groups who received the letter, those who were offered an incentive to register by email were four times more likely to register than those who received the letter ($p=0.001$). Those who received the request by email appear to be slightly more likely to register if offered an incentive, although this difference is not statistically significant. Among those who were not offered an incentive, those who received the email were five times more likely to register than those who received the letter ($p<0.0001$). There was no statistically significant difference between the two groups who received an incentive.

The lower part of the table shows the results for the re-allocated groups. Those who were in the "email, no incentive" group but where there was no email address were re-allocated to the "letter, no incentive" group. Thus, whilst the allocation to mode (letter or email) is no longer random, the allocation to incentive group is. For those who received a letter, being offered an incentive made a difference to whether the person registered ($p=0.001$). For those receiving an email, the offer of an incentive did not make a significant difference in registration rates.

[10] http://www.amazon.co.uk/gp/gc/
[11] http://store.apple.com/uk/browse/home/giftcards/itunes/gallery
[12] http://www.helpforheroes.org.uk/
[13] http://www.nspcc.org.uk/
[14] http://www.oxfam.org.uk/

This suggests that making it easier for sample members to participate, by clicking on an email link, encourages people to register. Where there is no email address for the sample member, offering a conditional incentive in a letter is more successful than not offering such an incentive.

Table 27: Number and proportion of registrations

	n	Number of registrations	% registered
Original			
Letter, no incentive	644	6	0.9
Email, no incentive	665	34	5.1
Letter, incentive	661	25	3.8
Email, incentive	661	37	5.6
Re-allocated			
Letter, no incentive	1056	11	1.0
Email, no incentive	246	29	11.8
Letter, incentive	1059	33	3.1
Email, incentive	253	29	11.5

Those, then, who claimed an incentive were a small group of adults and so we should be wary about drawing any conclusions from further analysis of this sub-sample. However, it is interesting to note that those who responded to the letter were more likely to choose the traditional £5 gift voucher (77.8% of letter redemptions) with just one person (5.6%) choosing an e-voucher. In the email incentive group, just 44% chose the £5 gift voucher and 24% (12 people) claimed an e-voucher. Of all redemptions, 13.2% gave their incentive to charity, whilst there were 10 people (20% of the email incentive group) who did not claim an incentive at all. Finally, those who registered with the website were more likely to give a full interview at the next wave (IP4) than those who did not (88.0% compared to 56.6%). Around three in ten adults who did not register were in a non-responding household at IP4, compared to just 4% of adults who registered. Of course, we cannot disentangle the suggestion that co-operative people who are more willing to participate would also be more willing to register with the study's website, and those who are more reluctant to 'tie themselves' to the study by registering are more likely to be non-responders at a future wave.

k. Mode effects in a self-completion instrument (Jonathan Burton)

Experiment proposers: Jonathan Burton, Peter Lynn

At Wave 3 of the main-stage *Understanding Society* study, the adult self-completion instrument changed from a paper questionnaire (used at waves 1 and 2) to a computer-assisted self-interviewing instrument (CASI). The CASI instrument required the interviewer to turn their laptop around so that the respondent could read and complete the questionnaire by themselves on the computer. The interviewer was available if the respondent had any queries or had difficulties using the computer. The rationale for changing the mode of the

self-completion instrument was that there would be a higher level of response using CASI since the respondent would complete it during the interview. With the paper self-completion the interviewer would distribute these to eligible adults once the household had been enumerated and collect them when they had been completed. Often the paper questionnaires were left in the household for the interviewer to pick up at a subsequent visit and on occasions the respondent was given a post-paid envelope to return the questionnaire. This procedure resulted in a number of questionnaires not being returned. Another reason for switching to CASI was that this would enable a more complex set of routing and filtering for questions, with the ability to feed-forward information from a previous interview or earlier in the same interview. On a paper self-completion the amount of routing which is feasible is quite limited without over-burdening the respondent. There are also cost and efficiency savings to be made for a CASI self-completion instrument; it is not necessary to print a large number of questionnaires and envelopes, there are no costs in scanning or coding the paper questionnaires and there is no requirement to then reconcile the questionnaire data with the self-completion data to ensure that a self-completion instrument is present for all those who were coded as having completed one.

At IP4 we had the opportunity to test the effect of switching to a CASI instrument, compared to administering a paper self-completion. Responding adults in a randomly allocated half of the households at IP4 were asked to complete a paper self-completion, as at previous waves of the IP. Adults in the other half of households were given the self-completion questions in a CASI module as part of their interview. The experiment aimed to test the assumptions about the response rate for the self-completion element. However, the experiment also aimed to look at the effect of the mode on measurement; whether the responses given on a paper self-completion were different from those given in a CASI mode. If there were measurement differences between the modes, we would investigate whether this was due to purely mode effects (the effect of completing the instrument on the computer compared to paper) or whether there were selection effects (those who completed the CASI instrument differed in some respect to those who completed the paper instrument). Since the experimental treatment was randomly allocated at the household level any selection effect would be due to differential non-response to the self-completion instrument.

Adults were eligible for the self-completion instrument if they responded to the individual questionnaire. Not all adults in responding households agreed to participate in the study, and so when looking at response rates to the self-completion instrument, the base is all adults who completed a full individual interview.

There were 1,113 adults who completed the individual interview and as part of that interview were asked to participate in the CASI instrument. Of these, 1,061 adults completed the CASI module, some with interviewer assistance, giving a response rate of 95.3%. A similar number of adults were asked to complete a paper self-completion questionnaire, 1,066. Self-completion data are available from 869 adults in this group, giving a response rate of 81.5%. A self-completion instrument delivered using CASI results in a higher unit response rate than one delivered using a paper self-completion questionnaire under these procedures ($p<0.001$).

We may also expect the mode of self-completion instrument to have an effect on item non-response. On a paper self-completion it is easier to leave a question blank than on the CASI instrument. Using the computer, if the respondent tries to pass through a question without giving a response, they are unable to proceed. The computer only responds if an answer is given. There are no explicit options for "don't know" or to refuse to answer a question, although the interviewer is able to show respondents how they can give these responses. The CASI instrument may also be perceived as being more 'private' that a paper self-completion. Once answered the response is in the computer and is not seen by the interviewer, whereas the paper questionnaire is given back to the interviewer, albeit in an envelope. At the end of the CASI section the interviewer enters their unique interviewer number which then 'locks' the CASI section so that it cannot be re-entered.

There were 69 question items in the self-completion modules which were common to both the CASI and paper self-completion groups. The mean level of item non-response (refusal, don't know, missing responses) for the CASI group was just 0.11 items per individual. For the paper group the level on item non-response was much higher, at 2.75 items per individual ($p<0.0001$). In the CASI group, 95.5% of respondents had no item missing values at all, compared to just 63.8% of the paper group. The proportion of respondents in the CASI group who had five or more missing items much lower than that of those in the paper group (0.56% compared to 18.3%). We can conclude, therefore, that the paper self-completion procedure resulted in higher unit non-response and higher item non-response.

As well as looking at the levels of response between the two modes, it is important to find out whether the substantive answers given in each mode were different. The mode of data collection may affect the responses given, as well as whether questions are answered or not. The design of the instrument, such as the way the question is placed on-screen or on the page, the way in which the response categories are aligned and the amount of space on the page or screen may affect the way a respondent answers. For example, at IP4 all the response categories in the CASI instrument were aligned vertically whilst in the paper self-completion there were some questions where the responses were aligned vertically, in the same layout as CASI, but in other questions the responses were printed horizontally.

There were statistically significant differences in responses to some of the questions between those who completed the paper self-completion instrument and those who completed the CASI. Simple t-tests suggest that those who completed a paper self-completion had a higher average number of friends (6.48) than those who completed using CASI (5.19) ($p<0.0001$). When answering this question in CASI, the respondent was just asked to "enter number". On the paper self-completion there was the instruction "Please enter number" and then two empty boxes. This may have given respondents to the paper self-completion an indication that the survey designers had the expectation that for some people this would require two digits. Almost twice as many respondents to the paper self-completion used two digits (23.1%) than in the CASI (12.9%).

Those who answered the paper self-completion appear to be in poorer health, were more likely to say that health limited them in moderate activities and in climbing several flights of stairs. They were also more likely to say that their physical health had meant that they had accomplished less than they would like. Those who answered the CASI were more likely than paper self-completers to say that pain had interfered with their work extremely, and less likely to say it had interfered a little bit. Those who completed a paper self-completion were more likely than CASI respondents to say that their physical health and/or emotional problems had interfered with their social activities all, some or a little bit of the time. These questions were all measured on a horizontal scale in the paper self-completion but vertical scale in CASI, although there is no consistency in response in that sometimes the paper self-completers favoured categories on the left and sometimes those on the right.

Whilst those who responded to the paper self-completion reported poorer health, in general, they were also more likely to say that they were satisfied with their health. They were also more likely to be satisfied with the income of their household and the amount of leisure time they had. However, there was no significant difference between modes on the level of satisfaction overall.

In a battery of questions on mental health which form the General Health Questionnaire (GHQ) those who completed a paper self-completion were more likely to give responses which were in the middle of a vertical scale, such as “no more than usual” or “same as usual”, to five of the twelve questions, in the other seven there was no significant differences between mode. Those who completed a paper self-completion were also more likely than those who responded using CASI to have trouble sleeping because they wake up in the middle of the night or because they cough or snore loudly. They were also more likely to report that they had taken medicine to help them sleep.

So it appears that the mode of self-completion instrument may affect measurement. However, although the allocation to treatment group was random there still may be some bias in the analyses due to sample selection. Since the paper self-completion had a lower response rate, it may be that the sample composition of those who completed the paper self-completion was systematically different to the sample who completed the CASI. These differences may lead to the differences we see in the measurement, rather than the mode itself. Using multi-nomial probit, and taking into consideration the complex sampling design of the survey (using svyset in Stata), we’re able to control for a number of socio-demographic variables. These analyses allow us to identify whether the mode of self-completion is associated with different measurement patterns, controlling for certain aspects of sample selection. We control for sex, age, marital status, labour market status and the number of children aged 14 or under for whom the respondent is responsible.

Thus, whilst a cross-tabulation suggests that those who respond to a paper self-completion are more likely that those using CASI to say that they are in good, fair or poor health, and less likely to report being in excellent or very good health, controlling for respondent characteristics suggests that there is no difference between modes in reporting fair or poor

health (compared to very good health). Those responding in CASI are still more likely to report excellent health than paper respondents, and vice versa for good health. The measurement effect of mode of self-completion is reduced once sample selection is taken into account, but does not disappear.

Controlling for respondent characteristics, those who responded to the paper self-completion were still more likely to say that their health limited them a lot (but no longer "a little") when it came to moderate activities, and that it limited them in climbing stairs. Without controls, paper self-completers were more likely to say that their physical health meant that they had accomplished less than they would like "all", "most", "some" and "a little bit" of the time compared to "none" of the time. Controlling for sample selection, only "some" of the time is reported more frequently for paper self-completers, with no mode effect for the other response categories. The suggested mode effect on the question about health interfering with social activities disappears when controls for respondent characteristics are used. The mode effects in five of the twelve GHQ questions, the questions about sleep and the satisfaction scales persist, controlling for respondent characteristics.

At IP4 we were also able to look at the time it took to finish the self-completion instrument. Whilst we always get timings for the computer-assisted instruments, we have not had timings for the paper self-completion. At IP4 we asked the respondent at the start of the paper self-completion to write in the time, and then at the end of the questionnaire we asked for the time again. Of the 869 paper self-completions, there was a start and finish time for 840 cases. In 11 cases, there was no timing information at all and for the other 18 cases there was either a start or end time, but not both. Of the 840 cases which had both a start and end time, 40 required some cleaning. This cleaning was predominantly required because one of the times was given using the 12-hour clock and the other using the 24-hour clock.

The mean average time to complete the paper instrument was 17.9 minutes, the median average was 13 minutes. However, there are a number of times which are not feasible. The CASI instrument was estimated to take a little under 11 minutes to complete. There are 59 cases of the paper self-completion taking less than five minutes and 109 cases taking longer than half an hour. Of those cases taking longer than half an hour, 40 were either 30, 60, 90 or 120 minutes, which may suggest some estimation or heaping around the half-hour. Trimming outliers is a subjective exercise, but if we analyse those self-completion instruments which longer than 5 minutes but less than half an hour, the mean average time to complete is 13.4 minutes, median 13 minutes. In CASI, just 44 cases took less than 5 minutes and there were no cases which took longer than half an hour. With the outliers trimmed, the CASI self-completion took a mean average of 10.2 minutes (median 9.4 minutes). It appears, then, that for cases where we have reliable timings and subject to our assumptions about outliers, the CASI was quicker to complete than the paper self-completion.

Table 28: Time of completion, paper and CASI

	Paper (minutes)	CASI (minutes)
All cases		
Mean	17.87	9.95
Median	13.00	9.18
Outliers removed		
Mean	13.45	10.23
Median	13.00	9.38

l. Implicit versus explicit requests for dates of events (Annette Jäckle)

Experiment proposers: Annette Jäckle, Noah Uhrig, Emanuela Sala, Fred Conrad

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 date 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.

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".

As already documented in Burton (2011), we expected to see the following effects:

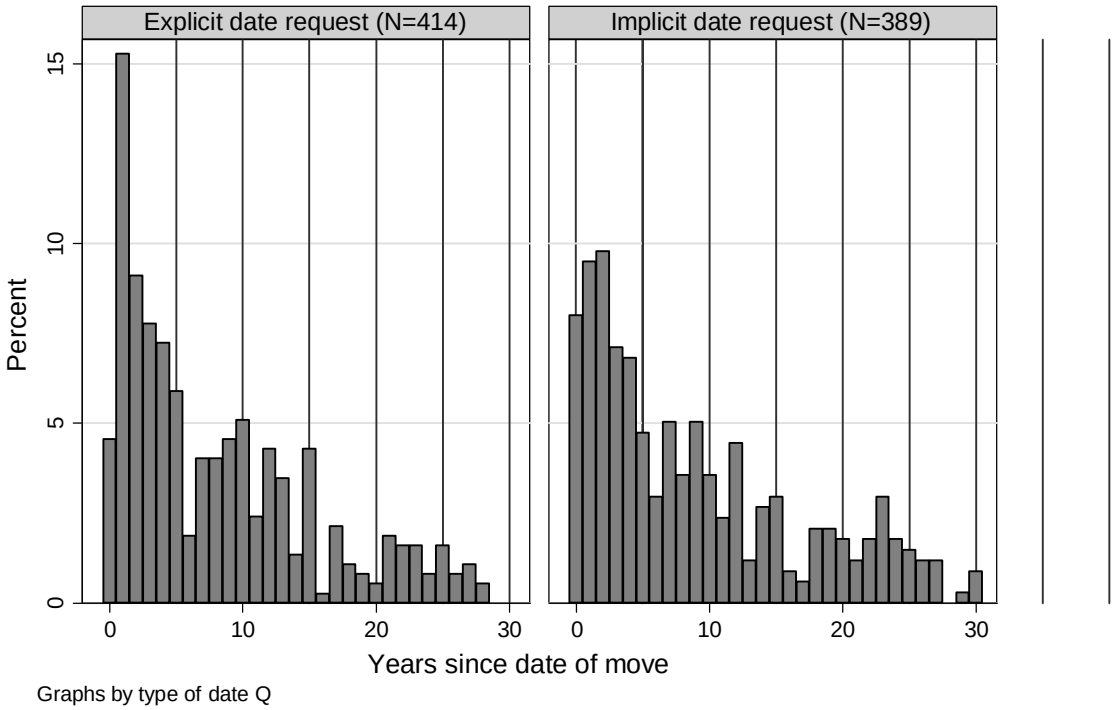
(1) The quality of date information is likely to be better with explicit rather than implicit requests. As a result we expect the elapsed duration since the event to be reported with more 'heaping' when the date request is implicit. That is, we expect more respondents to report durations (or dates that imply durations) of multiples of 5 years.
(2) We expect the wording of the question to affect which recall strategy respondents 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.

As the number of respondents reporting events in any one year is relatively small, the analyses we could do with the IP3 data alone were somewhat restricted. Pooling the IP3 and IP4 data we can further test the following expectations:

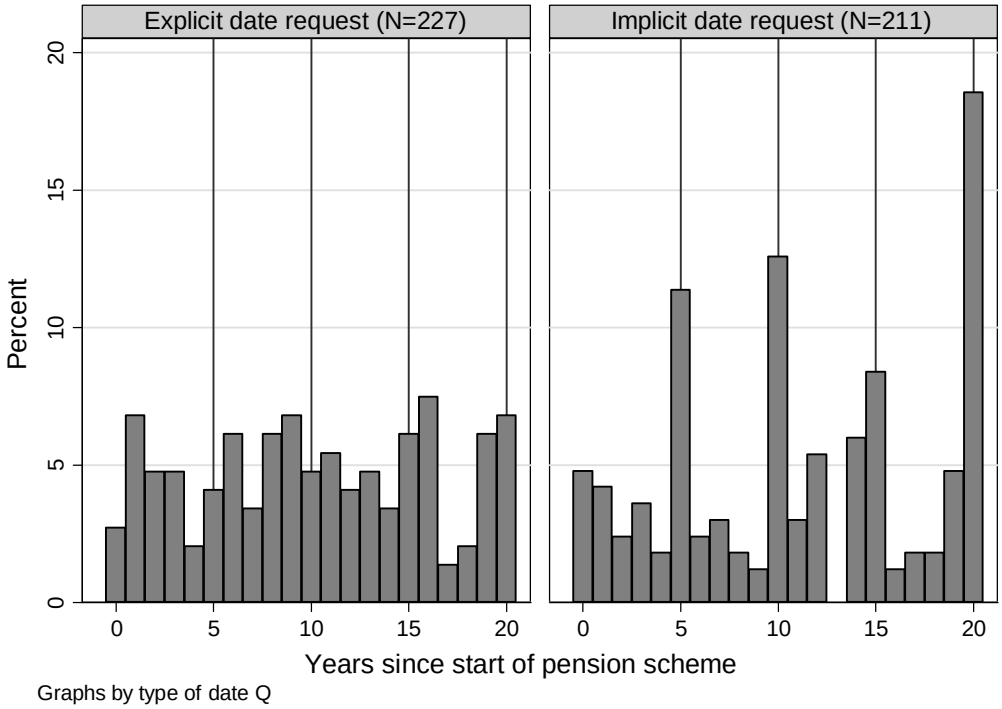
(3) We expect the recall strategy to affect the quality of the dates reported: We expect the probability of heaping to be lower for respondents who use direct recall strategies, than respondents who guessed or estimated dates of events.

(4) We expect certain groups of respondents to be more sensitive to the question instructions. For example respondents with lower ability, or of older age, may make the effort to directly recall the date of an event if it is explicitly requested, but may not make sufficient effort if the date request is implicit.

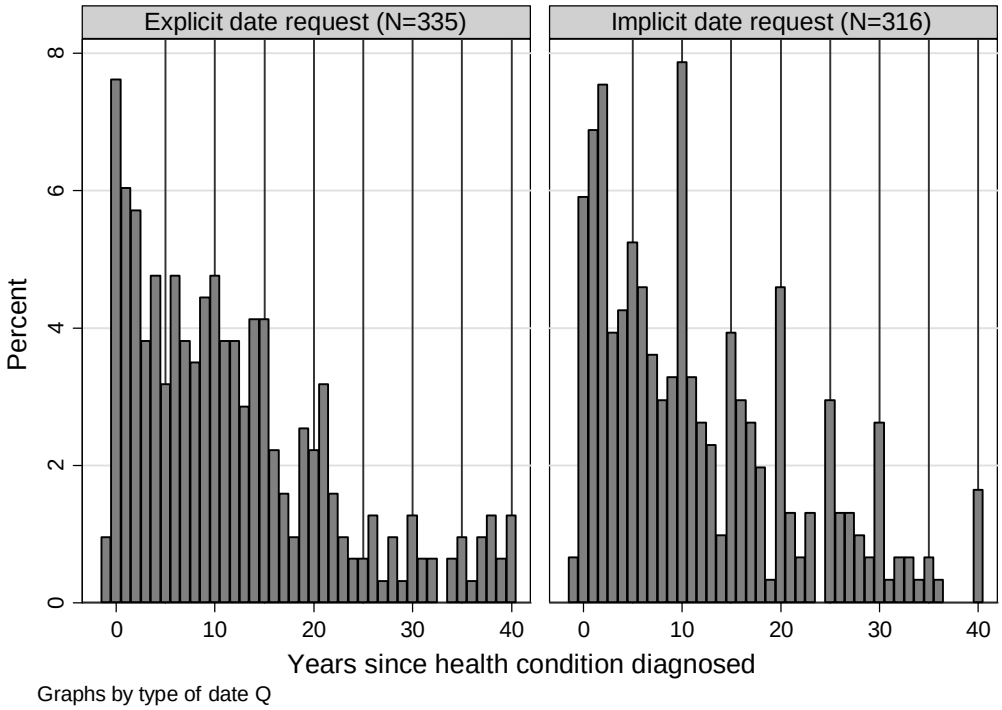
Graph 2: Years since residential move



Graph 3: Years since joined pension scheme



Graph 4: Years since health condition first diagnosed



Graphs 2 to 4 show the numbers of years since an event, separately for respondents who were asked the explicit question or the implicit version. Each graph pools responses from IP3 and IP4. Graphs 3 and 4 clearly show more heaping, that is, a larger proportion of respondents reporting durations (or dates that imply durations) that are multiples of 5 years. For dates of

residential moves (Graph 2) this effect is not apparent. Across both waves and the three date questions, the probability of a heaped response was 34.3% with the implicit date request, compared to 23.8% with the explicit date request (P-Value of Chi2 test adjusted for clustering in respondents =0.000, see Table 29).

Table 29: Probability of heaped response by question format

	Not heaped	Heaped (duration = multiple of 5 years)
Explicit request	76.2	23.8
Implicit request	65.7	34.3

Notes: P-value from Chi2 tests adjusted for clustering in respondents: P=0.000. N=1892 observations on N=1094 respondents.

As expected the retrieval strategy used by respondents to recall or reconstruct the date of an event is affected by the way the date is requested (Table 30). With the explicit date request 72.6% of respondents said they either knew the date, related it to other events or to characteristics of the event itself. With the implicit date request the proportion was only 60.8; a larger proportion of respondents in this group said they had guessed the date or related it to their own age (P-value of Chi2 test adjusted for clustering in respondents, P=0.000).

Table 30: Retrieval strategy by type of date request

	Guessed/related to own age	Knew/ related to other events/ characteristics of event
Explicit request	27.4	72.6
Implicit request	39.2	60.8

Notes: P-value of Chi2 test adjusted for clustering in respondents: P=0.000. N=1886 observations on N=1091 respondents.

Table 31 shows that the recall strategy used by a respondent is predictive of whether or not the respondent provides a heaped response. While 26.3% of respondents who used a direct recall strategy (i.e. knew the date, related it to the date of other events or to characteristics of the event) reported a heaped value, the proportion was higher (33.9%) for those who guessed or estimated the date of the event (P-value from Chi2 test adjusted for clustering in respondents, P=0.001).

Table 31: Probability of heaped response by recall strategy

	Not heaped	Heaped
Guessed/related to own age	66.1	33.9
Knew/ related to other events/ characteristics of event	73.7	26.3

Notes: P-value from Chi2 tests adjusted for clustering in respondents: P=0.001. N=1886 observations on N=1091 respondents.

To test whether certain types of respondents are more sensitive to the question format, we estimated the probability that the response is a multiple of 5 years (i.e. heaped), as a function of the question format, the respondent characteristic, and the interaction between respondent

characteristic and question format. The logit models were adjusted for clustering in respondents. The characteristics tested were sex, whether age 65+, and results from three tests of cognitive functioning (FAS test, working memory, and prospective memory. See Burton 2011 for a description of the cognitive tests.) The significance of interactions was tested using the –inteff– command in Stata (Norton, Wang, and Ai 2004).

None of the main effects were significant, meaning that there were no differences in the probability of providing a heaped response between men and women, older and younger respondents, and respondents with higher and lower scores on the cognitive tests.

None of the interaction effects were significant either, suggesting that there was no difference between different types of respondents in how their responses depended on the question format.

m. Ambiguous terms and concepts (Annette Jäckle)

Experiment proposers: Annette Jäckle, Noah Uhrig, Emanuela Sala, Fred Conrad

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/*Understanding Society* 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 in waves 3 and 4. Results from waves 2 and 3 are reported in Burton (2011). Currently no new results are available that include the wave 4 data.

n. Wording of dependent interviewing questions (Annette Jäckle)

Experiment proposers: Annette Jäckle, Noah Uhrig, Emanuela Sala, Fred Conrad

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. This experiment was designed to test whether the wording of dependent interviewing questions affects the measurement of change. Respondents were randomly allocated to two versions of dependent interviewing 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 experiment was first carried in wave 3 and repeated in wave 4 of the Innovation Panel. Burton (2011) contains a detailed description of the wave 3 experiment and results. Currently no new results are available that include the wave 4 data.

5 Plans for IP5

The Innovation Panel has been established as an important source of experimental data for methodological research. Over the first four 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 for interviewer-administered and self-completion instruments, different incentive strategies, the use of showcards, encouraging sample members to contact interviewers to arrange appointments 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, the context effect of asking for consent to data linkage and different ways of obtaining measures of household wealth. Since IP3, we have opened up the survey to researchers from around the world who are interested in experimenting with ways to improve survey measurement.

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 *Understanding Society* as well as to other research projects.

IP5 went into the field in May 2012 and had a mixed-mode design. One-third of the sample were allocated to face-to-face CAPI interviewers. The other two-thirds were allocated to a CAWI instrument, with a shift to face-to-face interviewers where the survey was not completed on-line. A competition for experiments was announced on 30th March and closed 2nd June 2011. There were 21 proposals submitted of which 15 were accepted. A summary of the experiments carried on IP5 are below. There will be more information on these in a future Working Paper.

i. Continued incentive experiment
ii. Continued mode of self-completion experiment (CASI versus paper)
iii. Branched vs unbranched rating scales, continuation of IP3 and IP4 experiment
iv. Panel conditioning and reliability of survey measures, continuation of IP1-IP4 experiment
v. Context effects in fertility decisions
vi. Mode preferences over time, replication of IP4 experiment
vii. Maximising whole-household response to a web questionnaire with incentives
viii. Effect of day of first contact in a web survey
ix. Effect of wording in dependent interviewing questions on the amount of change observed in panel data
x. Responses to ‘smiley faces’ versus text-based scales in the youth self-completion instrument
xi. Exploring methods to influence panel members response predisposition by manipulating the messages in the advance letter
xii. Altering the comparison group used when women answer questions about their satisfaction in different domains of life

xiii. Measuring partner satisfaction with the division of labour in the home using factorial vignettes
xiv. Assessing the feasibility of directly measuring household energy use
xv. Measuring subjective evaluations about the returns to schooling and the decision to go to university and testing the effect of an information treatment on actual decisions

A number of these experiments were audio-recorded in the face-to-face mode for additional analysis, with the consent of the respondent. The IP5 questionnaire is available on the *Understanding Society* website at:
http://data.understandingsociety.org.uk/documentation/innovation-panel/questionnaires#

REFERENCES

Cantor, D. (1989). Substantive Implications of Longitudinal Design Features: The National Crime Survey as a Case Study. Panel Surveys. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York John Wiley**:** 25-51.

Cantor, D. (2008). A Review and Summary of Studies on Panel Conditioning. Handbook of Longitudinal Research: Design, Measurement, and Analysis. S. Menard. London, Academic Press**:** 123-138.

Clausen, A. (1968). "Response validity: Vote report." Public Opinion Quarterly **32**: 588-606.

Cohen, S. B. and V. L. Burt (1985). "Data collection frequency effects in the National Medical Care Utilization and Expenditure Survey." Journal of Economic and Social Measurement **13**: 125-151.

Bailar, B. A. (1975). "The effects of rotation group bias in estimates from panel surveys." Journal of the American Statistical Association. **70**: 23-30.

Bartels, L. M. (1999). "Panel Effects in the American National Election Studies." Political Analysis **8**(1): 1-20.

Battaglia, M. P., E. R. Zell, et al. (1996). "Can participating in a panel sample introduce bias into trend estimates?" 1996 Proceedings of Survey Research Methods Section fo the American Statistical Association: 1010-1013.

Burton, J. (2011) "*Understanding Society* Innovation Panel Wave 3: Results from methodological experiments" Understanding Society Working Paper 2011-05, Colchester: University of Essex.

Ferber, R. (1953). "Obervations on a consumer panel operation." Journal of Marketing **17**: 246-259.

Ghangurde, P. D. (1982). "Rotation group bias in the LFS estimates." Survey Methodology **8**: 86-101.

Groves, R. M., Cialdini, R. B., and Couper, M. P. "Understanding the Decision to Participate in a Survey." Public Opinion Quarterly, 1992, 56, 475–495.

Kemsley, W. F. F. (1961). "The household Expenditure Enquiry of the Ministry of Labor." Applied Statistics **10**(3): 117-135.

Kraut, R. E. and J. G. McConahay (1973). "How being interveiwed affects voting: An experiment." Public Opinion Quarterly **37**(3): 398-406.

Kochanek, Kymn, Parvati Krishnamurty, Robert Michael, "The 'Early Bird' Field Innovation on a 30-year-old Longitudinal Survey". Paper presented at the 2nd Panel Survey Methods Workshop, Mannheim, Germany (5th-6th July 2010).

Mooney, H. (1962). Methodology in two California health surveys. US Public Health Service, Public Health Monograph No. 70. Washington, DC, US Government Printing Office.

Neter, J. and J. Waksberg (1964). "Conditioning effects fro repeated household interviews." Journal of Marketing **28**: 51-56.

Norton, E.C., H. Wang and C. Ai (2004). "Computing Interaction Effects and Standard Errors in Logit and Probit Models." The Stata Journal **4**:154-167.

Silberstein, A. R. and C. A. Jacobs (1989). Symptoms of repeated interivew effects in the Consumer Expenditure Interview Survey. Panel Surveys. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York, Wiley**:** 289-303.

Sturgis, P., N. Allum, et al. (2009). Attitudes Over Time: The Psychology of Panel Conditioning. Methodology of Longitudinal Surveys. P. Lynn. New York, John Wiley & Sons, Ltd**:** 113-126.

Turner, R. (1961). "Inter-Week Variations in Expenditure Recorded During a Two-Week Survey of Family Expenditure." Applied Statistics **10**(3): 136-146.

Uhrig, S.C.N. (2012) "Understanding panel conditioning: An examination of social desirability bias in self-reported height and weight in panel surveys using experimental data". Journal of Longitudinal and Life Course Studies, 3:1, pp 120-136.

Waterton, J. and D. Lievesley (1989). Evidence of conditioning effects in the British Social Attitudes Panel. Panel Surveys. D. Kasprzyk, G. Duncan, G. Kalton and M. P. Singh. New York, John Wiley & Sons**:** 319-339.

Wilson, S. E. and B. L. Howell (2005). "Do panel surveys make people sick? US arthritis trends in the Health and Retirement Survey." Social Science and Medicine **60**(11): 2623-2627.

Yalch, R. F. (1976). "Pre-election interview effects on voter turnout." Public Opinion Quarterly **40**: 331-336.