



**HAL**  
open science

## Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16

Antoine Coutrot, Tarek Al Baghal, Hugo J. Spiers

### ► To cite this version:

Antoine Coutrot, Tarek Al Baghal, Hugo J. Spiers. Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16. *Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data*, 2024 (11), pp.156-164, 2024. hal-04824916

**HAL Id: hal-04824916**

**<https://hal.science/hal-04824916v1>**

Submitted on 7 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Understanding Society**

**Working Paper Series**

**No. 2024 – 11**

**November 2024**

**Understanding Society Innovation Panel wave 16:  
Results from methodological experiments and new data**

Jim Vine, Edith Aguirre, Tarek Al Baghal, Michaela Benzeval, Jonathan Burton,  
Chris Butler, Heejung Chung, Mick Couper, Antoine Coutrot, Liam Delaney,  
Ceri Fowler, Annette Jäckle, Meena Kumari, Marion Lieutaud, Karen L. Mansfield,  
Lewis Mitchell, Violetta Parutis, John Payne, Gurleen Popli, Andrew K. Przybylski,  
Sharon Raj, Anita Ratcliffe, Adriaan R. Soetevent, Hugo Spiers,  
Gerard J. van den Berg, Lieke Voorintholt and Senhu Wang



## Non-technical summary

The *Understanding Society* survey includes what is known as an 'Innovation Panel' sample (IP). This sample is used to test different methods for conducting longitudinal surveys. The results from the Innovation Panel provide evidence about the best way to conduct a longitudinal survey, which is of relevance for survey practitioners working on other studies, as well as influencing decisions made about how to conduct *Understanding Society*. This paper reports the experiments, methodological tests, and other new data collection carried out at wave 16 of the Innovation Panel (IP16) from June to December 2023.

IP16 employed a mixed-mode design: some panel members were allocated to first be invited to a web version of the survey (with face-to-face as a follow-up mode), while others were invited to face-to-face first (with web as a follow-up mode); in both cases, telephone interviews were used as the third attempted mode. IP16 also continued ongoing experiments on the impact of incentives.

As with prior waves, several other methodological experiments were included in the survey. Survey methods experiments / evaluations were conducted on: how to measure worries about climate change; how to measure depression; the use of e-vouchers for unconditional incentives; how to encourage responses to the youth survey; how to obtain information about child development measures; the effects of incentives on respondents' participation in a study using a game-like app; and what effect a government logo on envelopes sent to existing panel members would have on response rates.

In addition, IP16 was used to try out new survey content. New content was carried related to: what constitutes an interesting effect size when measuring people's psychological state; people's expectations of gender discrimination related to work; the extent and nature of the use of domestic workers, and how people find such workers; numeracy and the long-term future; the extent to which working flexibly is stigmatised; judging the passage of time; and associations between cognitive reflection abilities and politically motivated reasoning.

Other new data was gathered as part of IP16 and is presented in this paper relating to: child development measures from children's 'red book' records; and spatial navigation ability data from a game-like app.

## Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data

Jim Vine<sup>1</sup>, Edith Aguirre<sup>1</sup>, Tarek Al Baghal<sup>1</sup>, Michaela Benzeval<sup>1</sup>, Jonathan Burton<sup>1</sup>, Chris Butler<sup>2</sup>, Heejung Chung<sup>3</sup>, Mick Couper<sup>4</sup>, Antoine Coutrot<sup>5</sup>, Liam Delaney<sup>6</sup>, Ceri Fowler<sup>7</sup>, Annette Jäckle<sup>1</sup>, Meena Kumari<sup>1</sup>, Marion Lieutaud<sup>6</sup>, Karen L. Mansfield<sup>8</sup>, Lewis Mitchell<sup>1</sup>, Violetta Parutis<sup>1</sup>, John Payne<sup>1</sup>, Gurleen Popli<sup>9</sup>, Andrew K. Przybylski<sup>8</sup>, Sharon Raj<sup>6</sup>, Anita Ratcliffe<sup>9</sup>, Adriaan R. Soetevent<sup>10</sup>, Hugo Spiers<sup>11</sup>, Gerard J. van den Berg<sup>12</sup>, Lieke Voorintholt<sup>10</sup> and Senhu Wang<sup>13</sup>

<sup>1</sup> Institute for Social and Economic Research, University of Essex, <sup>2</sup> University of Antwerp, <sup>3</sup> King's College London, <sup>4</sup> University of Michigan, <sup>5</sup> Centre National de la Recherche Scientifique, <sup>6</sup> London School of Economics and Political Science, <sup>7</sup> University of Oxford, <sup>8</sup> Oxford Internet Institute, University of Oxford, <sup>9</sup> Department of Economics, University of Sheffield, <sup>10</sup> University of Groningen, <sup>11</sup> University College London, <sup>12</sup> University of Groningen, University Medical Center Groningen, and IFAU Uppsala, <sup>13</sup> National University of Singapore

**Abstract:** This paper presents some preliminary findings from wave 16 of the Innovation Panel (IP16) of *Understanding Society: The UK Household Longitudinal Study*. *Understanding Society* is a major panel survey in the UK. In June 2023, the sixteenth wave of the Innovation Panel went into the field. IP16 used a mixed-mode design, using on-line interviews, face-to-face interviews, and telephone interviews. This paper describes the design of IP16, the experiments and other studies carried, new datasets produced, and the preliminary findings from early analyses of the data.

**Keywords:** longitudinal, survey methodology, experimental design, question wording, questionnaire design, dataset.

JEL classification: C80, C81, C83

**Acknowledgements:** *Understanding Society* is an initiative funded by the Economic and Social Research Council (ESRC) and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey

delivery by NatCen Social Research and Verian (formerly known as Kantar Public). Wave 16 of the Innovation Panel was funded by ESRC grant ES/T002611/1. The research data are distributed by the UK Data Service.

**Contact:** Jim Vine ([jim.vine@essex.ac.uk](mailto:jim.vine@essex.ac.uk)), Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

# Contents

<b>1. Introduction.....</b>	<b>9</b>
1.1 References .....	10
<b>2. Understanding Society: The UK Household Longitudinal Study .....</b>	<b>11</b>
2.1 References .....	13
<b>3. Innovation Panel wave 16 .....</b>	<b>14</b>
3.1 Call for studies .....	14
3.2 Sample .....	15
3.3 Questionnaire design.....	15
3.4 Response outcomes.....	15
3.5 References .....	16
<b>4. Studies carried in IP16.....</b>	<b>17</b>
<b>5. Measuring worries about climate change: The effect of a subtle wording change ....</b>	<b>19</b>
5.1 Introduction .....	19
5.2 Methods.....	20
5.3 Results.....	21
5.4 Conclusion.....	23
5.5 References .....	23
<b>6. The prevalence of depression in Understanding Society .....</b>	<b>25</b>
6.1 Experimental design .....	26
6.2 Preliminary results .....	27
6.3 Conclusion and next steps .....	27
6.4 References .....	28
<b>7. Using e-vouchers for unconditional incentives.....</b>	<b>29</b>
7.1 Introduction .....	29
7.2 Methods.....	30
7.3 Results.....	31
7.4 Conclusion.....	33
7.5 References .....	34

<b>8. Youth survey experiments.....</b>	<b>36</b>
8.1 Introduction .....	36
8.2 Methods.....	37
8.3 Results.....	38
8.4 Conclusion.....	40
8.5 References .....	40
8.6 Appendix: IP16 youth leaflets:.....	42
<b>9. Obtaining measurement data from the ‘red book’ .....</b>	<b>46</b>
9.1 Design.....	46
9.2 Results.....	48
9.3 Conclusion.....	49
9.4 References .....	50
<b>10. The effect of incentives on the willingness to participate in a gamified app study ....</b>	<b>51</b>
10.1 Motivation .....	51
10.2 The Sea Hero Quest app .....	52
10.3 Research questions.....	52
10.4 Study design.....	53
10.5 Results.....	54
10.6 Summary .....	58
10.7 References .....	59
10.8 Appendix.....	61
<b>11. Use of a government logo .....</b>	<b>62</b>
11.1 Introduction .....	62
11.2 Methods.....	63
11.3 Results.....	63
11.4 Conclusions .....	65
11.5 References .....	66
<b>12. Establishing a databank of smallest effect sizes of interest (SESOI) for prominent psychological constructs.....</b>	<b>67</b>
12.1 Introduction .....	67

12.2 Methods.....	68
12.3 Results.....	69
12.4 Discussion and Conclusions .....	73
12.5 References .....	74
<b>13. Anticipated gender discrimination in the labour market .....</b>	<b>76</b>
13.1 Summary and key findings.....	76
13.2 Motivation .....	76
13.3 Question included in the Innovation Panel (IP).....	78
13.4 Sample and preliminary analysis of responses.....	78
13.5 Analysis .....	80
13.6 Demographic characteristics .....	81
13.7 Job characteristics.....	84
13.8 Discussion and conclusion .....	86
13.9 References .....	87
<b>14. Domestic workers and platforms.....</b>	<b>89</b>
14.1 Introduction and context.....	89
14.2 Methods and data.....	90
14.3 Analysis .....	90
14.4 References .....	94
<b>15. Numeracy and the long-term future .....</b>	<b>96</b>
15.1 Introduction .....	96
15.2 Methods.....	97
15.3 Results.....	99
15.4 Conclusion.....	102
15.5 References .....	103
<b>16. Flexibility stigma in the UK post-pandemic : intersectionality of stigma across gender and parental lines .....</b>	<b>105</b>
16.1 Introduction .....	105
16.2 Methods.....	109
16.3 Results.....	114



16.4 Conclusion.....	119
16.5 References .....	120
<b>17. The relationship between employment status, passage of time judgements and wellbeing.....</b>	<b>124</b>
17.1 Introduction .....	124
17.2 Methods.....	125
17.3 Results.....	127
17.4 Conclusion.....	131
17.5 References .....	131
<b>18. Cognitive reflection and politically motivated reasoning.....</b>	<b>133</b>
18.1 Introduction .....	133
18.2 Methods.....	135
18.3 Initial results.....	138
18.4 References .....	145
<b>19. Child development measures from the ‘red book’ .....</b>	<b>148</b>
19.1 Introduction .....	148
19.2 Methods.....	148
19.3 Data.....	149
19.4 References .....	155
<b>20. Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16 ...</b>	<b>156</b>
20.1 Introduction .....	156
20.2 Methods.....	157
20.3 Data.....	161
20.4 References .....	163

# 1. Introduction

This working paper presents early findings from wave 16 of the Innovation Panel (IP16) of *Understanding Society: The UK Household Longitudinal Study*.

*Understanding Society* is a major panel survey for the UK. It has a large sample size (40,000 households at wave 1), has benefited from an ethnic minority boost sample, and has included the collection of biomarker and genetic/epigenetic data. The first fourteen waves of data collection on the main sample have been completed, and the fifteenth and sixteenth waves are currently in the field. The data from the first fourteen waves of the main samples are available from the UK Data Service (University of Essex, Institute for Social and Economic Research 2024a).

One of the features of *Understanding Society* 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). The Innovation Panel is used for methodological testing and experimentation that would not be feasible on the main sample. It has been used to test different fieldwork designs, new survey questions and new ways of asking existing questions. The design of the Innovation Panel, fieldwork and outcomes, and the content of the survey at each wave are documented in the User Guide (Institute for Social and Economic Research 2024).

This paper describes the innovations carried in IP16, the sixteenth wave of the Innovation Panel. After briefly summarising some features of the design of *Understanding Society* in general and the Innovation Panel in particular, this Working Paper has chapters describing studies carried in IP16. IP16 differed from earlier waves of the Innovation Panel in explicitly being open to non-experimental studies that would be suited to data collection on the Innovation Panel. Consequently, the scope of this Working Paper is somewhat broader than previous Working Papers in the sequence, reporting on these other studies as well as the experimental studies conducted within IP16.

Data for all completed waves of the Innovation Panel (including IP16) are available from the UK Data Service (University of Essex, Institute for Social and Economic Research 2024b). Past Working Papers covering experiments carried out in previous waves of the Innovation Panel are available from the *Understanding Society* website.<sup>1</sup>

## **1.1 References**

Institute for Social and Economic Research (2024). *Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-16, User Guide*.

Colchester: University of Essex,

<https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide>.

University of Essex, Institute for Social and Economic Research. (2024a). *Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009*.

[data collection]. *19th Edition*. UK Data Service. SN: 6614, [DOI:](https://doi.org/10.5255/UKDA-SN-6614-20)

<https://doi.org/10.5255/UKDA-SN-6614-20>

University of Essex, Institute for Social and Economic Research. (2024b). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. *13th Edition*.

UK Data Service. SN: 6849, [DOI: https://doi.org/10.5255/UKDA-SN-6849-16](https://doi.org/10.5255/UKDA-SN-6849-16).

---

<sup>1</sup> <https://www.understandingsociety.ac.uk/research/publications/type/understanding-society-working-paper-series/>

## 2. Understanding Society: The UK Household Longitudinal Study

*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. As a longitudinal panel survey, it aims to collect data on the same households year after year. For the main sample, data has now been collected over 14 waves, providing many opportunities for researchers to investigate issues in ways that would not be possible with cross-sectional data. In fact, because the former British Household Panel Survey (BHPS) sample was incorporated into the *Understanding Society* sample at wave 2, the data available spans decades for some sample members.

The study is managed by the Executive Team (ET), based at the Institute for Social and Economic Research (ISER) at the University of Essex and includes topic experts from a number of institutions. The fieldwork and delivery of the survey data for the first five waves of the main sample were undertaken by the National Centre Social Research (NatCen). Since wave 6, Verian (formerly known as Kantar Public) has been the lead contractor.

*Understanding Society* aims to be the largest survey of its kind in the world. The sample covers the whole of the UK, including Northern Ireland and the Highlands and Islands of Scotland. *Understanding Society* provides high quality, longitudinal survey data for academic and policy research across different disciplines. The use of geo-coded linked data enables greater research on neighbourhood and area effects, whilst the collection of biomarkers and physical measurements has increased the utility of the survey for health analysts.

The design of the main *Understanding Society* study is similar to other household panel studies 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 questionnaire. Sample members are

interviewed annually, although the fieldwork for each wave of the main sample is spread over two years: for each wave, the sample is split into 24 monthly batches with one issued to the field every month.

*Understanding Society* is a mixed-mode study. For several years, the majority of fieldwork has been conducted via the web, with some respondents still completing face-to-face or by telephone. For those completing face-to-face there is also a self-complete section, where interviewers pass the laptop to the respondents to complete questions viewed as being more sensitive.

In between each wave of data collection, sample members are sent short reports of early findings from the survey, and a change-of-address card, to allow them to inform ISER of any change in their address and contact details. Before each sample month is issued to field, each adult is sent a letter which informs them about the new wave of a survey, enclosing a change-of-address card. For sample members who responded at the previous wave, the letter also includes an unconditional incentive (gift voucher); previous-wave non-responders are offered the same incentive, conditional on response.

Households are enumerated, whether online or by an interviewer, seeking information on who is in the household including any new entrants and the location of anyone who has moved from the household. The first person in the household to complete the survey is asked to complete the household enumeration; one member of each household (ideally a bill-payer) is also asked to answer the household questionnaire (if no one in the household has completed it yet); and all eligible members of the household are asked to complete an individual questionnaire. New entrants are eligible for inclusion in the household. Those who move within the UK are traced and interviewed at their new address. People living with sample members are also temporarily eligible for interview. More information about the sampling design of *Understanding Society* is available in Lynn (2009).

## **2.1**    *References*

Lynn, P. (2009). Sample Design for Understanding Society. *Understanding Society Working Paper Series No. 2009 – 01*. Colchester, Essex: Institute for Social and Economic Research, University of Essex.

<https://www.understandingsociety.ac.uk/research/publications/514007>

### **3. Innovation Panel wave 16**

The design of the Innovation Panel's fieldwork protocols, sample following rules, and questionnaire content are modelled on the main *Understanding Society* survey. It has around 1,500 households, and the first wave of interviews was conducted in 2008. Unlike the main sample, Innovation Panel waves are all conducted within one year, with the fieldwork period typically running from around spring until the end of the year. Also unlike the main sample, Northern Ireland and areas north of the Caledonian Canal are excluded. Refreshment samples of around 500 respondent households each have been added in waves 4, 7, 10, 11, and 14 of the Innovation Panel.

IP16 used a mixed mode design, with households allocated to face-to-face-first or web-first designs. Sample members who had not completed in the mode they were initially issued to after a number of weeks were followed-up and invited to take part using the other mode. Web-first sample members who had not completed their interview online in the first five weeks were issued to CAPI interviewers to follow-up while those issued to interviewers were invited to complete online after 10 weeks. Towards the end of fieldwork, interviewers tried to contact adults who had not yet responded in either mode, in order to interview them via a telephone (CATI) interview.

Fieldwork took place between 21<sup>st</sup> June and 8<sup>th</sup> December 2023.

#### **3.1 *Call for studies***

As in previous waves, at IP16 the Innovation Panel was again open for researchers outside the scientific team of *Understanding Society* to propose studies. Whilst previous waves' calls had focused on survey methods experiments/evaluations, for IP16 the call explicitly also solicited proposals for new survey questions to be carried in the questionnaire. A public call for proposals was made and 18 proposals were received with 10 being accepted. Proposals were reviewed by subject-matter reviewers, members of the *Understanding Society* executive team, and the fieldwork agency. In addition to the proposals accepted through

the public call, 5 further studies were included by the executive team in order to develop methodology for the main *Understanding Society* survey. One further experiment that is core to the fieldwork design of the study — the mixed-mode design — was continued from past waves.

### **3.2 Sample**

Households from six sample origins were issued at IP16: the original sample from IP1 and refreshment samples issued at IP4, IP7, IP10, IP11, and IP14. There were 583 original sample households, 257 IP4 refreshment sample households, 312 IP7 refreshment sample households, 219 IP10 refreshment sample households, 379 IP11 refreshment sample households and 755 IP14 refreshment sample households issued (Institute for Social and Economic Research 2024, table 13.39).

### **3.3 Questionnaire design**

The questionnaire at IP16 followed the standard format used in the previous Innovation Panels as well as the main survey of *Understanding Society*. The Innovation Panel questionnaire each wave is based on the most recent main questionnaire, but with rotating content modules omitted in order to make room for the experimental content and other studies being run. The questionnaires used at IP16 are available from the *Understanding Society* website.<sup>2</sup>

### **3.4 Response outcomes**

The issued sample at IP16 wave comprised 2,505 households. There were 1,748 interviewed households from the continuing samples, for a 70% overall household response rate.

---

<sup>2</sup> <https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires>



(Institute for Social and Economic Research 2024, table 13.39) Within these households, 2,825 people were interviewed, for an individual response rate of 57% (Institute for Social and Economic Research 2024, table 13.40).

Further details of the response rates are reported in the Innovation Panel User Guide (Institute for Social and Economic Research 2024, tables 13.39, 13.40, 13.41, 13.42, and 13.43).

### **Longitudinal response outcomes**

The individual re-interview rate is an important outcome in a longitudinal panel study since many analyses require pairs of observations to measure change. A re-interview rate at a given wave is calculated as the percentage of units who were surveyed at the initial wave and remain eligible who respond at the later wave. For those in the original sample, the percentage is predicated on response at IP1, while for the various refreshment samples their respective initial waves were IP4, IP7, IP10, IP11, and IP14.

As with any longitudinal study, there has been attrition over time, decreasing the overall numbers for each sample. At IP16, the response rates for the samples ranged from 25% of initial sample members (for whom IP16 was their 16<sup>th</sup> wave) being re-interviewed up to 68% of the IP14 refreshment sample (for whom IP16 was their third wave), with the IP4 to IP11 refreshment samples all having response rates between 36% and 42%. Full longitudinal re-interview rates are reported in the Innovation Panel User Guide (Institute for Social and Economic Research 2024, table 13.43).

## **3.5 *References***

Institute for Social and Economic Research (2024). Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-16, User Guide.

Colchester: University of Essex,

<https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide>.

## 4. Studies carried in IP16

The following chapters contain summaries of the studies fielded in IP16, describing the design of each study and some initial results from early analysis of the data. The analyses in this working paper were based on a preliminary dataset that contained all cases but did not have weights or derived variables. The authors and proposers of each study are given in the respective chapters.

In the IP16 competition, proposers specified whether their proposed study related to survey methods experiments/evaluations or new survey questions. The latter did not preclude the proposal having an experimental component.

Two data description chapters are included, describing novel datasets that were generated as a result of work conducted as part of IP16. In addition to the usual expected practice of citing the Innovation Panel dataset as a whole, authors who make use of either of these datasets are encouraged to cite the respective data description chapters too.

	Chapter
<i>Survey methods experiments/evaluations</i>	
Measuring worries about climate change: The effect of a subtle wording change	5
The prevalence of depression in Understanding Society	6
Using e-vouchers for unconditional incentives	7
Youth survey experiments	8
Obtaining measurement data from the 'red book'	9
The effect of incentives on the willingness to participate in a gamified app study	10
Use of a government logo	11
<i>New survey questions</i>	
Establishing a databank of smallest effect sizes of interest (SESOI) for prominent psychological constructs	12
Anticipated gender discrimination in the labour market	13
Domestic workers and platforms	14

Numeracy and the long-term future	15
Flexibility stigma in the UK post-pandemic : intersectionality of stigma across gender and parental lines	16
The relationship between employment status, passage of time judgements and wellbeing	17
Cognitive reflection and politically motivated reasoning	18
<i>Data descriptions</i>	
Child development measures from the 'red book'	19
Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16	20

In addition to the experiments and new content described in this working paper, a further project was specified for IP16. This was a project intended to provide improved guidance on how to use proxy nominations for panel members moving into a care home. No panel members were identified as moving into a care home at IP16; consequently, the new content was not used and there is no data on the performance of the content to present in this paper.

## 5. Measuring worries about climate change: The effect of a subtle wording change

*Lieke Voorintholt, University of Groningen*

*Adriaan R. Soetevent, University of Groningen*

*Gerard J. van den Berg, University of Groningen, University Medical Center Groningen, and IFAU Uppsala*

### 5.1 Introduction

Climate change is posing extreme risks to the world and effective mitigation strategies are necessary to limit future damages (IPCC 2023). The successful implementation of mitigation policies largely depends on public support and acceptance. To accurately measure people's attitudes toward climate change and related policies, we need robust measurement tools. Correct measurement not only better informs policymakers of the potential receptivity to their initiatives but also advances research into the factors driving these attitudes, thereby guiding the design of effective behavioural interventions.

This experiment concerns the measurement of people's *worries* about climate change. The Understanding Society main survey includes a recurring question about climate change worries, appearing in waves 4 (2012-2014) and 10 (2018-2020) (University of Essex, Institute for Social and Economic Research 2020). The question on worries about the effects of climate change and the corresponding module on environmental attitudes is unique for similar longitudinal studies such as PSID, SHARE, and HILDA. Respondents are asked how much they agree (on a 5-point scale) with the following statement: "The effects of climate change are too far in the future to really worry me."

The statement's reference to (future) effects of climate change makes it suitable to study intergenerational valuation. Since the effects of climate change get worse over time, they will mostly impact future generations and cause large intergenerational inequities (Thiery et al. 2021). For this reason (elderly) people's worries about climate change partly capture their valuation of the welfare of future generations. Multiple studies show a negative age

trend for various environmental constructs and the literature attempts to explain this (Andor, Schmidt, & Sommer 2018; Geys, Heggedal, & Sørensen 2020; Milfont et al. 2021). Nonetheless, conclusions are still indeterminate, and this topic requires more research using suitable (survey) data, for example, the question in the Understanding Society survey mentioned above.

However, one concern regarding this question is the potentially flexible interpretation of the statement. Different people might read the question with an emphasis on different aspects, e.g. *too far*, *really worry*, *worry me*. Specifically, the reference to personal worries by the addition of ‘me’ differs from the few household surveys that have included a climate change concern question, such as SOEP and the European Social Survey, and might limit the potential to study intergenerational valuation with this instrument. With our survey experiment we aim to determine whether including this reference leads to a different (range of) response compared to a climate change worry question without this reference.

## **5.2    *Methods***

We analyse the difference between two versions of this question with data from wave 16 of the Understanding Society Innovation Panel (University of Essex, Institute for Social and Economic Research 2024). Half of the sample was asked how much they agree with the original statement “The effects of climate change are too far in the future to really worry me” (control group), whereas the other half was asked about their agreement with “The effects of climate change are too far in the future to really worry about” (treatment group). Answer options range from 1 (totally agree) to 5 (totally disagree).

Our outcome variable of interest is the climate change worry question for any of the two treatments. We first visually inspect the distribution of answers to both questions. Following this, we regress worry level on treatment exposure to inspect the mean difference between the two question variations. We complement this basic model with more elaborate models that also include interaction effects of treatment with various demographic variables. This

allows us to investigate whether there exists heterogeneity in the effect of treatment exposure.

### 5.3 Results

Our current results are based on 2705 observations of respondents with completed observations for all our outcome and control variables. This estimation sample is split into a control (N = 1359) and treatment (N = 1346) group.

**Figure 5-1 Distribution of climate change worries by treatment**

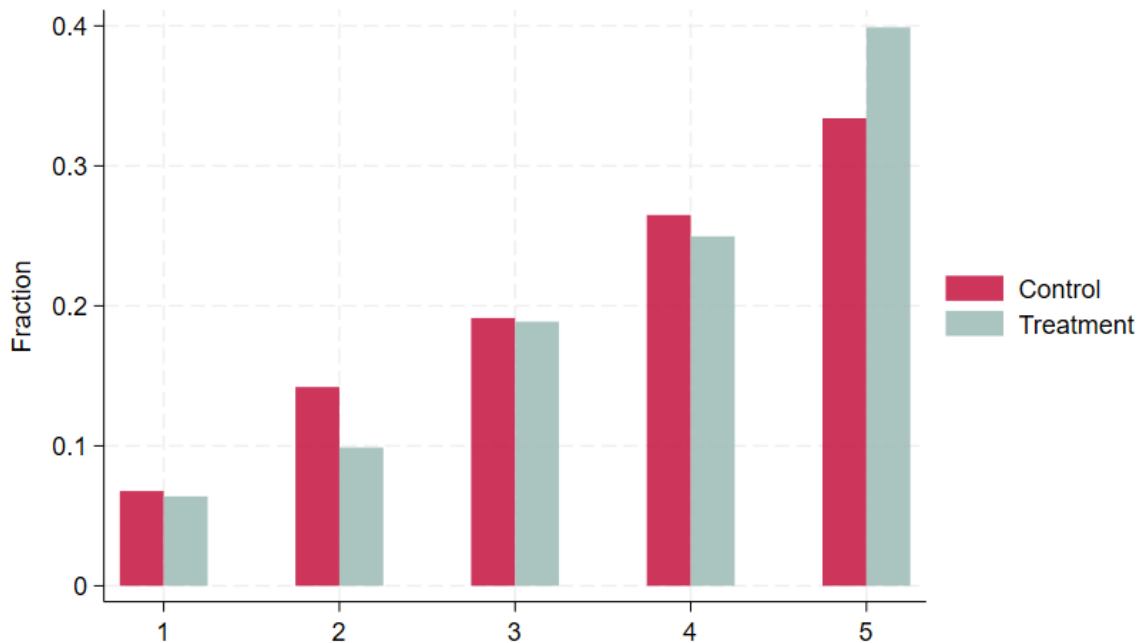


Figure 5-1 shows the distribution of agreement with the statement “The effects of climate change are too far in the future to really worry me” (control group) or “The effects of climate change are too far in the future to really worry about” (treatment group). For both groups, the most common response is total disagreement with this statement, implying a worry score of five. In the treatment group, a higher proportion of respondents choose this answer compared to the control group. For most other answer options, both groups appear to perform similarly. An exception is the ‘tend to agree’ option (2), which is more often selected by respondents in the control group than in the treatment group.

**Table 5-1 OLS estimates of wording effect interacted with relevant demographics**

	(1)	(2)	(3)	(4)	(5)
	CC worries	CC worries	CC worries	CC worries	CC worries
Treat	0.17*** (0.05)	0.18*** (0.05)	0.19* (0.09)	0.21** (0.07)	0.16 (0.09)
Treat # age 36-55			-0.12 (0.13)		
Treat # age 56-65			-0.11 (0.14)		
Treat # age 66+			0.17 (0.13)		
Treat # female				-0.05 (0.10)	
Treat # parent					0.02 (0.10)
Constant	3.66*** (0.03)	3.25*** (0.10)	3.25*** (0.11)	3.24*** (0.10)	3.26*** (0.10)
Controls	No	Yes	Yes	Yes	Yes
Observations	2705	2705	2705	2705	2705

Notes: Robust standard errors in parentheses. Controls consist of the dummy variables: female, parent, working, married, very good or excellent health (above median), fair or poor health (below median), and the approximate quartile age groups 36-55 years, 56-66 years, and 66+ years (omitted age group is 35-years). The standard deviation of our outcome is approximately 1.25. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 5-1 translates the distributions from Figure 5-1 into testable differences in the average level of climate change worries. The estimates indicate that the worry level is significantly higher for respondents in the treatment group than in the control group by on average around 0.18, approximately 15% of one standard deviation. This result implies that the ‘worry me’ ending of the statement on climate change worries leads to somewhat more agreement than the ‘worry about’ ending. A possible explanation would be that some respondents interpret the control statement as worries about the effect of climate change on them personally, whereas the treatment statement is interpreted more generally or even normatively, namely whether climate change is something that ‘one’ should worry about.

In columns (3), (4), and (5) we interact the treatment effect with various relevant demographics. The estimates show that our main treatment effect remains relatively stable and always positive, across all models. The interactions with age group, sex, and parenthood

are all insignificant. As a result, we cannot yet identify whether the treatment effect is driven by a specific subset of respondents.

## **5.4 Conclusion**

The results of this study show that exact wording matters for measuring worries about climate change. Specifically, we find that the question originally used in the Understanding Society Main Survey (University of Essex, Institute for Social and Economic Research 2020) leads to a somewhat lower average worry score than that of an altered version without the reference to personal worries (by replacing the word ‘me’ with ‘about’).

We posit that this effect is due to a potentially different interpretation of the control and treatment statement. Figure 5-1 shows that a small subset of respondents tend to agree (option 2) with the ‘worry me’ variant of the statement, whereas they might have completely disagreed (option 5) with the ‘worry about’ variant. This suggests that the moderate magnitude of the positive treatment effect is driven by a limited number of respondents who are strongly influenced by the precise wording in their interpretation of the statement. We will conduct additional analyses to further explore this proposed explanation and try to identify the affected subgroup.

Our results are relevant for the design of environmental modules in future surveys by highlighting the importance of exact wording. Future research could help uncover the mechanisms behind such wording effects and determine if they can be employed to increase support for climate change mitigation.

## **5.5 References**

Andor, M. A., Schmidt, C. M., & Sommer, S. (2018). Climate change, population ageing and public spending: Evidence on individual preferences. *Ecological Economics*, 151, 173–183.



- Geys, B., Heggedal, T.-R., & Sørensen, R. J. (2020). Popular support for environmental protection: A life-cycle perspective. *British Journal of Political Science*, 1–8.
- IPCC. (2023). *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team, H. Lee and J. Romero, Eds.). IPCC.  
<https://doi.org/10.59327/IPCC/AR6-9789291691647>
- Milfont, T. L., Zubielevitch, E., Milojev, P., & Sibley, C. G. (2021). Ten-year panel data confirm generation gap but climate beliefs increase at similar rates across ages. *Nature Communications*, 4038(12).
- Thiery, W., Lange, S., Rogelj, J., Schleussner, C. F., Gudmundsson, L., Seneviratne, S. I., Andrijevic, M., Frieler, K., Emanuel, K., Geiger, T., Bresch, D. N., Zhao, F., Willner, S. N., Büchner, M., Volkholz, J., Bauer, N., Chang, J., Ciais, P., Dury, M., ... Wada, Y. (2021). Intergenerational inequities in exposure to climate extremes. *Science*, 374(6564), 158–160.
- University of Essex, Institute for Social and Economic Research. (2020). *Understanding Society: Waves 1-10, 2009-2019 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. 13th Edition. UK Data Service. SN: 6614, <http://doi.org/10.5255/UKDA-SN-6614-14>
- University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## 6. The prevalence of depression in Understanding Society

*Meena Kumari, Institute for Social and Economic Research, University of Essex*

*Jonathan Burton, Institute for Social and Economic Research, University of Essex*

When Understanding Society was initiated, the common assumption was that the prevalence of mental health conditions was 'one in four'. This statistic was the basis for a number of campaigns and public health initiatives at the time (Ginn and Horder 2012). However, this statistic is contested and more recent evidence suggests that the prevalence of mental disorder varies from 1% for psychosis to one in six for less severe common mental disorder (Bebbington and McManus 2019).

In Wave 1 (2009-2011), participants were asked whether a doctor had ever diagnosed clinical depression within a checklist of a number of conditions (University of Essex, Institute for Social and Economic Research 2023). Clinical depression has a prevalence of 6% in Wave 1, which is clearly considerably lower than the 'one in four' expectation. There may be a number of reasons for this perceived underestimate. The wording of the question is focussed on depression, which does not capture all mental health disorders. Further, the focus on 'clinical depression' may have been confusing to participants. Thus, in Wave 10 (2018-2020) of the study, the question was altered to capture a wider array of common mental disorders uncoupled from 'clinical' conditions and participants were asked whether a doctor had ever diagnosed 'An emotional, nervous or psychiatric problem' followed by a list of common mental health conditions. In Wave 13 (2021-2023) participants were asked this question but were presented with an expanded list of mental health conditions to capture additional mental health conditions relevant to the COVID-19 pandemic.

It is not clear how changes in these questions have impacted the reporting of mental health conditions in Understanding Society. For example, the reporting of depression changed from a direct question (Wave 1-9) to being embedded in short (Waves 10-12) and long (Wave 13) list of conditions routed through a general question on mental health problems.

## 6.1 *Experimental design*

Here we randomly allocated households to three equal sized groups. Adult participants were administered one of three versions of the conditions questionnaire: (i) the Waves 1-9 version; (ii) the Waves 10-12 version; and (iii) the Wave 13 version. For face-to-face (CAPI) and telephone (CATI) interviews, the interviewer read out the question. Web participants completed online, as a self-completion questionnaire (University of Essex, Institute for Social and Economic Research 2024).

The question starts with: “Has a doctor or other health professional ever told you that you have any of these conditions.” The participant is then read or shown a list of conditions. In CAPI interviews, there is a showcard that the participant can look at when answering. The participant in interviewer-administered surveys can say the condition, or just the number on the showcard.

To measure the prevalence of depression, we examine the proportion of people in each group who gave or selected a response option which identified depression.

In the first (Wave 1-9) version, this was “Clinical depression” (question `HCONDA`, code 17), selected from a list of 20 items covering both physical and mental health.

In the second (Waves 10-12) version, the participant would first have to select “An emotional, nervous or psychiatric problem” (question `HCONDB`, code 22) from a list of 20 items covering both physical and mental. They would then be asked “What type of emotional, nervous or psychiatric problem was that?” (question `MHEALTHTYPB`) and we use those who answered “Depression” (code 2) or “Bipolar disorder or manic depression” (code 4) from a list of 7 possible responses.

In the third version (Wave 13), we identify as having depression those that selected “Bipolar disorder (or ‘manic depression’)” (code 4), “Depression” (code 2), or “Nervous breakdown” (code 13) from a list of 18 possible responses, all related to mental health conditions. In this version, this question about mental health conditions (`MHCONDC`) was asked of all

respondents in that allocation, before also asking them a question about physical health conditions (HCOND, 25 items).

## 6.2 Preliminary results

The reported prevalence of depression derived from the waves 1-9 and waves 10-12 version of question are similar to each other in this experiment. Further, they are similar to the reported prevalence apparent in waves 1-9 in the main stage of Understanding Society. Interestingly, an expanded list of mental health conditions is associated with greater reporting of depression than with the earlier version of the questionnaire (see Table 6-1, below).

**Table 6-1 Reported prevalence of depression, by question version**

	<b>Version 1</b> <b>Waves 1-9</b>	<b>Version 2</b> <b>Waves 10-12</b>	<b>Version 3</b> <b>Wave 13</b>	<b>Total</b>
Depression	52 5.2%	46 4.9%	167 18.4%	265 9.3%
Total	996	937	910	2,843

## 6.3 Conclusion and next steps

We conclude that there was an under-reporting of mental health conditions in the earliest waves of Understanding Society. This is because we focussed on depression alone. If we include all of the emotional, nervous or psychiatric problems in the second version, we would identify a greater prevalence of these problems with 7.3% of participants reporting positively. Focussing on depression, the reported prevalence using the waves 1-9 version was similar to the Waves 10-12 version of the questionnaire. However, the prevalence was much lower than the Wave 13 version. This might be for two reasons: the Waves 10-12 version of the question had two stages, with the prevalence of depression ascertained in the second step. In the wave 13 version, participants were asked about mental health conditions in a separate question, not embedded within a long list of conditions. Further

analyses are required to understand these differences in reporting and the impact on associations typically investigated by the research community.

## **6.4**     *References*

Bebbington PE, McManus S. (2020). Revisiting the one in four: the prevalence of psychiatric disorder in the population of England 2000-2014. *Br J Psychiatry*. 2020 Jan;216(1):55-57. doi: 10.1192/bjp.2019.196. PMID: 31685036; PMCID: PMC7557610.

Ginn S, Horder J. (2012). "One in four" with a mental health problem: the anatomy of a statistic. *BMJ*. 2012 Feb 22;344:e1302. doi: 10.1136/bmj.e1302. PMID: 22362250.

University of Essex, Institute for Social and Economic Research. (2023). *Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. 18th Edition. UK Data Service. SN: 6614, <http://doi.org/10.5255/UKDA-SN-6614-19>.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## 7. Using e-vouchers for unconditional incentives

*Violetta Parutis, Institute for Social and Economic Research, University of Essex*  
*Jonathan Burton, Institute for Social and Economic Research, University of Essex*

### 7.1 Introduction

The positive impact of incentives on response rates in cross-sectional and longitudinal surveys has been well documented (e.g. Toepoel, 2012; Laurie & Lynn, 2009; Cabrera-Álvarez & Lynn, 2023). There is also research evidence to favour the use of unconditional survey incentives compared to conditional ones (Messer & Dillman, 2011; Toepoel, 2012), although evidence in this area is mixed (Booker et al. 2011). Some research findings also show that monetary incentives are more effective than non-monetary ones (e.g. Laurie and Lynn 2009; Singer & Ye, 2013). Incentives play an especially important role in encouraging response in web and mixed mode surveys which typically have lower response rates compared to face-to-face surveys (e.g. Bianchi et al. 2016). Some findings from Understanding Society: The UK Household Longitudinal Study (UKHLS) suggest that increasing incentive amounts to mixed-mode allocated survey participants can increase response rates to the level of those observed in the face-to-face designs (Jäckle et al. 2015, Bianchi et al. 2016, Gaia 2017). The role of incentives in mixed mode longitudinal surveys still remains understudied. UKHLS Innovation Panel experiments have provided some useful insights into this topic over the years. For example, Jäckle and Lynn (2008) found that unconditional incentives were more effective in reducing attrition in a panel of young people in the UK than conditional ones. Gaia (2017) compared the effect of a £30 unconditional incentive to a £10 unconditional plus a £20 conditional incentive and found that although differences in response between the two conditions were small, the unconditional incentive yielded higher response rates in subsequent waves.

Given that the push-to-web design is gaining popularity (Dillman 2017) due to its potential to save on fieldwork costs, and the importance of retaining sample members in longitudinal

surveys, it is especially important to gain a better understanding about how incentives on mixed mode longitudinal surveys affect response and therefore how to use them effectively.

Understanding Society offers £20 conditional incentives to previous-wave non-responders and unconditional incentives to previous-wave responders in the form of Love2Shop gift cards. Those who are invited to complete their survey online are also offered a £10 bonus conditional on completion of their interview in the first five weeks of fieldwork. In addition to gift cards, survey participants who qualify for a conditional incentive can choose their incentive in the form of an e-voucher or a charity donation. Unconditional incentives only come in the form of gift cards and are sent with advance letters. The lack of choice here is mainly due to the technical challenges related to devising a system able to deal with unconditional e-vouchers, but also partly with a concern that asking sample members to do something to get their voucher, even going online to claim it, may be perceived as being no longer “unconditional”. The experiment described below sets out to test the feasibility and acceptability of offering unconditional incentives in the form of e-vouchers (University of Essex, Institute for Social and Economic Research 2024). We hypothesise that using e-vouchers gives participants a greater choice in the type of incentive they receive and may therefore have a positive effect on their motivation to participate in the survey. Rather than being restricted to physical Love2Shop gift-cards that can only be spent at physical stores, e-incentives will let participants choose vouchers that can be used online or a donation to charity, alongside the Love2Shop gift-card.

## **7.2    *Methods***

As noted above, in Understanding Society unconditional incentives are used for previous-wave responding adults, so the response figures in the tables below are re-interview rates. Adult sample members eligible for unconditional incentive were randomly allocated at the household level to two groups. One group received the Love2Shop gift-card, as standard, in their advance letter. The second group received their advance letter in which they were

invited to go online and claim their unconditional e-voucher. The link to access the e-incentive was added to the advanced letters for the experimental group. Those in the experimental condition also received one of three versions of the ‘motivational message’ regarding why e-incentives are better than gift cards.

- 1 = no message (control group)
- 2 = eco-friendly
- 3 = easier to use than gift cards
- 4 = both eco-friendly and easier to use than gift cards

### 7.3 Results

The results show a significant difference in survey individual response between the two experimental groups: 86% in the control groups vs 81% in the experimental (e-incentives) group. This suggests that survey participants were more likely to respond when offered the standard incentives, i.e. a gift card, rather than the e-incentive. No significant difference in response was found in terms of age or sex.

**Table 7-1 Individual survey outcome by incentives experimental group**

Individual survey outcome	Experimental groups		Total
	Control (gift cards)	E-incentives	
Not interviewed	194 (14%)	249 (19%)	440 (16%)
Interviewed	1,183 (86%)	1,071 (81%)	2,254 (84%)

Pearson chi2(1) = 10.3 Pr = 0.001

Further analyses of response by the mode of issue (web vs face-to-face) reveal that the differences between the experimental groups are only significant for those who were issued web-first, but not those who were issued CAPI-first.



**Table 7-2 Individual survey outcome among those issued by incentives experimental group and mode of first issue**

Individual survey outcome	Experimental groups		Total
	Control (gift cards)	E-incentives	
<b>Issued CAWI-first<sup>1</sup></b>			
Not interviewed	97 (10%)	150 (17%)	247 (13%)
Interviewed	840 (90%)	744 (83%)	1,584 (87%)
<b>Issued CAPI-first<sup>2</sup></b>			
Not interviewed	97 (22%)	96 (23%)	193 (22%)
Interviewed	343 (78%)	327 (77%)	670 (78%)

<sup>1</sup> Pearson chi2(1) = 16.2 Pr < 0.001

<sup>2</sup> Pearson chi2(1) = 0.05 Pr = 0.819

This suggests that the form of the incentive is more relevant for survey members who complete the survey online compared to those who complete with an interviewer.

Table 7-3, below, indicates that there was no significant difference in response between the incentive groups which were sent different motivational messages about e-incentives.

**Table 7-3 Individual survey outcome by e-incentives motivational messages experimental group**

Individual survey outcome	Motivational messages experimental groups				Total
	Control (no message)	eco-friendly	easier to use	eco-friendly and easier to use	
Not interviewed	54 (16.2%)	70 (20.1%)	63 (19.6%)	59 (18.8%)	246 (18.7%)
Interviewed	279 (83.8%)	278 (79.9%)	259 (80.4%)	255 (81.2%)	1,071 (81.3%)

Pearson chi2(1) = 1.9 Pr = 0.578

## **7.4 Conclusion**

Given the scarcity of evidence on how different forms of survey incentives influence survey response, especially in longitudinal surveys, we explored the effects of two forms of unconditional incentives, gift cards vs e-vouchers. In contrast to our original expectations, we found that survey participation was higher among those who were offered the standard gift card compared to those who were offered e-incentives this wave. This suggests that in longitudinal surveys participants may become used to receiving the same form of incentive each year and do not like when it is changed. Because this experiment focused on unconditional incentives, the results suggest that participants prefer to receive their incentive in the post rather than have to claim them, which requires additional effort and time on their part. The clear preference for the physical gift cards among the survey participants was also noted by our study helpline team who received an increased number of calls from the sample members, most of them (50+) asking to replace their e-vouchers with standard gift cards. Unfortunately, we were not able to receive information on how many of those who were offered e-vouchers actually redeemed them, which would allow to shed some more light on this discussion about the usefulness of e-incentives to our survey participants. Similarly, we do not know how many of those who intended to claim their e-incentive might have encountered technical problems to do so. Our participant helpline has received some calls from participants asking to re-send the email with the instructions and links allowing them to claim their incentives. Finally, it will be interesting to see how this experiment affected survey participation next wave of those who were assigned to the e-incentive group this wave. In addition to the findings related to survey response, the experiment also revealed some technical challenges related to being able to offer sample members unconditional e-incentives which they needed to be able to claim at the point of the being invited to the interview, and which needed to be accumulated if the participant did not claim the unconditional incentive before their interview and also became eligible for a conditional (early bird) incentive when completing the interview.

## 7.5 *References*

- Bianchi, A., Biffignandi, S., & Lynn, P. (2016). Web-CAPI sequential mixed mode design in a longitudinal survey: effects on participation rates, sample composition and costs. Understanding Society Working Paper 2016 – 08, *Colchester: University of Essex*.
- Cabrera-Álvarez, P., & Lynn, P. (2023). Short-term impact of increasing the value of unconditional and conditional incentives in Understanding Society. Understanding Society Working Paper Series 2023-08, *Colchester: University of Essex*.
- Dillman, D.A. (2017). The promise and challenge of pushing respondents to the Web in mixed-mode surveys. *Survey Methodology*, 43 (1), 3-30
- Gaia, A. (2017). The Effect of Respondent Incentives on Panel Attrition in a Sequential Mixed-mode Design. Understanding Society Working Papers, 2017-3, *Colchester: University of Essex*.
- Jäckle, A., & Lynn, P. (2008). Respondent incentives in a multi-mode panel survey: Cumulative effects on nonresponse and bias. *Survey Methodology*, 34(1), 105–117.
- Laurie, H., & Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 205–233). John Wiley & Sons, Ltd.
- Jäckle, A., Lynn, P., & Burton, J. (2015). Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Time and Unit Non-Response. *Survey Research Methods*, 9(1), 57-70.
- Messer, B. L., & Dillman, D. A. (2011). Surveying the General Public over the Internet Using Address-Based Sampling and Mail Contact Procedures. *Public Opinion Quarterly*, 75(3), 429–457. <https://doi.org/10.1093/poq/nfr021>
- Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *Annals of the American Academy of Political and Social Science*, 645, 30. <https://doi.org/10.1177/0002716212458082>

Toepoel, V. (2012). Effects of Incentives in Surveys. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 209–223). Springer New York.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## 8. Youth survey experiments

*Violetta Parutis, Institute for Social and Economic Research, University of Essex*  
*Jonathan Burton, Institute for Social and Economic Research, University of Essex*

### 8.1 Introduction

As survey response rates have been declining world-wide (Luiten et al. 2020), young people seem to be especially difficult to engage. To-date there has been very limited research exploring what factors play a significant role in motivating survey response among young people specifically (Parutis 2023). Such a knowledge gap especially affects longitudinal surveys since over the years children and young people in such surveys grow up and transition into adult participants. Limited research evidence suggests that those who take part as children are more likely to stay in the survey as adults (Parutis 2023).

Although it is reasonable to expect that factors affecting young people's survey participation differ from those that mitigate response in adults, traditional survey methods literature suggests that the most relevant factors include modes, devices, study materials and incentives (Flanagan et al. 2015). Previous research comparing paper and online surveys for children and young people found that overall response was better on web-based surveys (Scott-Johnson et al., 2010, cited in Flanagan et al. 2015, p.10). However, others found that youth response was lower on push-to-web surveys compared to face-to-face surveys (Kantar Public 2023). There is some evidence to suggest that young people prefer to complete surveys on mobile devices (Denny et al. 2008, cited in Flanagan 2015, p.9), although more recent studies found that this applies more to older children rather than younger ones, since the latter tend to use computers and laptops for this purpose and complete surveys under parental supervision (Kantar Public 2023). Since it is a standard practice to send survey invitations to children via their parents to ensure parental consent, the role of parents in young people's survey participation also merits further investigation (ONS, 2023). Finally, while the role of incentives in adult surveys has been widely

investigated, including in Understanding Society (e.g. Cabrera-Álvarez & Lynn 2023), there is very little evidence on how incentives work in youth surveys.

In Understanding Society (UKHLS), the majority of adult sample members are invited to complete the survey online, with non-responders followed up by an interviewer. 10-15-year-olds are invited to complete the survey using paper booklets which are posted to responsible adults in the household once they complete their adult survey online. If an interviewer visits a household with a 10-15-year-old, they administer a paper booklet directly to the young person, subject to parental consent. Although we have experimented with online youth surveys a couple of times on the Innovation Panel in the past, the take up of the online option has been extremely low. We therefore continue to explore factors that might help to motivate young people to respond, especially online. In this study specifically we looked at the effect of an additional conditional incentive on youth response rates and the effect of different versions of the information leaflet directed at either the young person or their parent (University of Essex, Institute for Social and Economic Research 2024).

## **8.2    *Methods***

Children in households in which the household grid was completed online or by telephone were sent the invite to the youth survey by post. This included a paper questionnaire, a covering letter with a QR code and URL and access code to complete online (addressed to the parent), an unconditional incentive (£10), and a leaflet. In households where the household grid was completed in person, the interviewer handed over the youth self-completion paper questionnaire, covering letter (with QR code, URL, and access code to complete online), and unconditional voucher.

We experimented with two elements: the use of an additional conditional incentive; and the type of leaflet enclosed with the questionnaire.

In a random half of households, the covering letter mentioned that if the child completed the survey online or returned the questionnaire by post, they would be sent an additional £5 gift card.

The mailing also included a leaflet which gave more information about the survey, what it was used for, and why it was important for researchers. There were two versions of the leaflet, where the content and design were either targeted at the young person, or targeted towards the parent. The children's leaflet aimed to be more visually attractive and included information about the incentive, the type of questions, confidentiality of the responses as well as some motivational messages about why it is important to take part. The parents' leaflet was much more text-heavy and included information about why we ask children to complete their own survey, why we ask some sensitive questions about drugs and alcohol, and confidentiality of the responses (please see the Appendix for screenshots of the two leaflets). Households were randomly allocated to the type of leaflet.

For practical reasons, children in the households issued to an interviewer had the offer of an additional conditional incentive and the leaflet included in the first reminder (sent from the office), rather than in the initial invite.

### **8.3 Results**

In total around half (n=130) of eligible<sup>3</sup> young people completed the survey. 47 responded online, while 83 completed and returned paper booklets. Overall, the results show the effect of the additional conditional incentive on survey response was not statistically significant.

---

<sup>3</sup> A 10-15-year old sample member becomes eligible for a youth survey as soon as a responsible adult in the household completes their adult survey. In unproductive households where no individual adult surveys are completed, youths are not invited to take part.

**Table 8-1 Youth survey outcome by incentive experimental group**

Youth survey outcome	Experimental groups		Total
	Additional conditional £5	Control	
Not interviewed	69 (53.5%)	58 (45.3%)	127 (49.4%)
Interviewed	60 (46.5%)	70 (54.7%)	130 (50.6%)

Pearson  $\chi^2(1) = 1.7$  Pr = 0.2

However, in households where the interviews were conducted in-person, and the interviewer handed over the youth self-completion pack, the additional conditional incentive was significant at the  $p < 0.1$  level. However, this was in the opposite direction to what we had expected, with the control group having a higher response rate, than the group promised the additional £5 incentive.

**Table 8-2 Youth survey outcome by incentive experimental group in households where the grid was completed with an interviewer**

Youth survey outcome	Experimental groups		Total
	Additional conditional £5	Control	
Not interviewed	25 (59.5%)	22 (40%)	47 (48.5%)
Interviewed	17 (40.5%)	33 (60%)	50 (50.6%)

Pearson  $\chi^2(1) = 3.6$  Pr = 0.06

Furthermore, the results show the effect of the leaflet variation on youth response was not statistically significant.



**Table 8-3 Youth survey outcome by leaflet experimental group**

Individual survey outcome	Experimental groups		Total
	Child-targeted leaflet	Parent-targeted leaflet	
Not interviewed	71 (53%)	56 (45.5%)	127 (49.4%)
Interviewed	63 (47%)	67 (54.5%)	130 (50.6%)

Pearson  $\chi^2(1) = 1.4$  Pr = 0.2

## **8.4 Conclusion**

Given the scarcity of evidence on what motivates young people to respond in surveys, especially in the longitudinal context, we experimented with offering the youths an additional conditional incentive and two different versions of the information leaflet. Overall, results of the analysis showed no statistically significant effect of any of these interventions, although they suggested that additional conditional incentive may be a significant discouraging factor in households visited by an interviewer – although it is not clear what the mechanism of this effect would be. Future research (Innovation Panel 17) in this area will experiment with a re-designed (more engaging) online youth survey and different versions of the survey invitation mailings.

## **8.5 References**

- Cabrera-Álvarez, P., & Lynn, P. (2023), Short-term impact of increasing the value of unconditional and conditional incentives in Understanding Society. Understanding Society Working Paper Series 2023-08.
- Flanagan, S., Greenfield, S., Coad, J. & Neilson, S. (2015). An exploration of the data collection methods utilised with children, teenagers and young people (CTYPs). BMC Research Notes. 8 (1): 61.
- Kantar Public (2023). Research on Transforming the Crime Survey for England and Wales Work package B: Exploring the use of an online survey to measure crime among

children and young people. [Accessed online

[\[https://kantarpublish.com/articles/transforming-the-crime-survey-for-england-and-wales\]](https://kantarpublish.com/articles/transforming-the-crime-survey-for-england-and-wales)

Luiten, A., Hox, J., & de Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics*, 36, 469-487.

ONS (2023) CCSEW Online Pilot 1: Findings and recommendations report. Power Point presentation. [shared by the research team at ONS].

Parutis, V. (2023). Recruiting and retaining young people in longitudinal surveys: lessons from Understanding Society: UK Household Longitudinal Study. *Understanding Society Working Paper Series 2023-15*.

<https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2023-15.pdf>

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## 8.6 Appendix: IP16 youth leaflets:

### 1. Leaflet targeted at the child



**My life is important.**

We want to hear all about you, so people (like the government and charities) can make better decisions about things that affect young people.

PAGE 2  
(BACK OF FRONT COVER)  
210 x 146mm

**"I think young people should be heard."**

We ask questions on some serious topics, like bullying, drinking, and how safe you feel where you live. The survey covers things that you will have to deal with, but we want to hear about your experiences.

PAGE 4, 208 x 146mm

**Thank you...**

for being part of Understanding Society!

Scan the QR code to find out what life is like now and what's changed over time.

**My life is super busy!**

Friends, school, family, hobbies - there's lots going on. Your survey helps us see how young people balance their lives and what they think is most important.

PAGE 7, 208 x 146mm

**"I'm part of something special."**

You're part of the biggest household survey in the UK - and one of the largest in the world!

There are other young people completing similar surveys in the USA, Japan, Germany, South Korea, Australia,....

PAGE 9  
210 x 146mm

**My life is super busy!**

Friends, school, family, hobbies - there's lots going on. Your survey helps us see how young people balance their lives and what they think is most important.

TOTAL FLAT SIZE:  
620 x 440mm

## 2. Leaflet targeted at the parents

**Children are really important to Understanding Society, as their survey answers help us understand more about issues that directly affect them.**

Every participant in Understanding Society represents thousands of other people living in the UK. We ask everyone in the household to complete an interview, to help us see how families are changing over time.

Very few studies in the world ask children and teenagers about their experiences over several years. Ours is the **only study in the UK** that captures children's experiences in this way. This makes the information that your child gives to Understanding Society **extremely important** - and we are very grateful when children and their families take part.

**Please encourage your child to complete their survey because it is helping thousands of other children in the UK when they take part.**



**What if you have questions?**

If you are worried about any of the questions we ask your children, you can contact our participant liaison team.

[www.understandingsociety.ac.uk/participants](http://www.understandingsociety.ac.uk/participants)

You can contact our participant liaison team by email: [contact@understandingsociety.ac.uk](mailto:contact@understandingsociety.ac.uk)

By Freephone: 0800 252 853

Or write to us: **FREEPOST BR0X-KEKJ-JGKS, Understanding Society, University of Essex, Wivenhoe Park, Colchester, CO4 3SD.**




**Your child**  
and Understanding Society

Why do we ask children to take part?





UK  
Economic and Social  
Research Council  
IP16 Youth Leukos (Parent)



iSER  
INTEGRATED SURVEILLANCE RESEARCH ENVIRONMENT



University of Essex



## Why do we give children their own survey?

The questions we ask your child cover the key areas of their life and include things like family life, friends, school, social media, health and leisure. Some of these topics may include questions about more serious issues, such as bullying, drinking alcohol or vaping.

Giving children their own survey allows us to tailor the questions to their experiences and to ask about things that happen at home, at school or in the street, that parents might not experience.

Children have told us that they like having their own special survey, as it shows that their experiences are important and they have a vital role in Understanding Society.

## Why do we allow you to complete your survey in private?

Experience tells us that completing surveys on their own allows children to feel more open when they answer the questions. They can give their answer, rather than the 'right' answer.

Please do read the questions in your child's survey before you give it to them. You may want to discuss with them any topics that you think may confuse or worry them. Or you may want to talk to them afterwards about anything that completing the survey has made them think about.



## Are your child's survey answers confidential?

Yes. Your child's personal details such as name, address and date of birth are removed from their survey answers, so that they cannot be identified. Their details are never made available to anyone outside the Study team.

Just like your survey, we want you to treat your child's answers to their survey as confidential. When they have completed their questionnaire, they should seal it in the envelope and hand it back to the interviewer, or give it to you to post back.



## Why do we ask questions on sensitive topics?

Topics such as drinking alcohol, smoking and vaping, drug use, bullying, and internet safety may not apply to your child directly, but these are issues that can impact children within this age range. By collecting this information, we are able to track what's happening with children across the UK.

We are careful to include only topics that your child will have learnt about at school, as part of the National Curriculum. The survey questions are developed with help from specialist researchers who work with children.

If you are worried about any of the questions, you can contact our Participant Liaison Team who can provide more information on why we ask these kinds of questions.



## Why do we include questions on...

### Alcohol?

In your child's survey we ask whether they have friends who drink alcohol, if they drink alcohol themselves and, if they do, how often.

Information from children's surveys in Understanding Society has helped researchers look at what makes young people less likely to drink or smoke. They discovered that children who know about the harmful effects of drinking and smoking were less likely to do either. Also, that children who reported higher happiness levels were less likely to drink or smoke.

### Bullying?

We ask children if they experience bullying at school, at home, or online. We also ask if they have bullied someone.

Bullying is a big issue for children in the UK - and there are many reasons why children bully each other. Researchers have used Understanding Society to look at how children's behaviour has changed since the Covid pandemic. They found that the pandemic school closures have increased children's stress levels, and this has made them more likely to bully and fight. This research has been shared with organisations that support teachers and with the UK government.

### Wellbeing?

Children in Understanding Society are asked how they feel about different aspects of their life. We are interested in how they feel about their appearance, their family, their friends, school, and life as a whole.

Information on children's wellbeing is vital for planning support services for children and their families. Each year The Children's Society produces The Good Childhood Report on the mental health of children in the UK, using information from children in Understanding Society. This report shows that children's happiness levels have been falling over time for things like school and appearance, but that most children remain happy with their family life.



If completing their survey makes your child want to talk to someone outside their family about the issues raised, we include a number of helplines that they can call or text.

## 9. Obtaining measurement data from the ‘red book’

*Michaela Benzeval, Institute for Social and Economic Research, University of Essex*

*Edith Aguirre, Institute for Social and Economic Research, University of Essex*

*Tarek Al Baghal, Institute for Social and Economic Research, University of Essex*

*Lewis Mitchell, Institute for Social and Economic Research, University of Essex*

Understanding Society is a unique resource for research on childhood, often having data on both parents before children are born, siblings, and then annual data on the family with detailed data on the children at key milestones. To create further research opportunities, having early objective data on child development would be very valuable. However, directly and regularly measuring physical health is invasive and time consuming for families, and costly for studies. In wave 16 of the Innovation Panel (IP16), therefore, we investigated whether it was possible to collect such data from the NHS ‘Personal Child Health Record’ – administrative records parents hold for their children, known as the ‘red book’. The red book is a paper booklet where health visitors and parents record key development milestones for children under the age of 2 years. At the time of developing this experiment, some health authorities were beginning to move the red book to a digital record.

Our aims in this project were to investigate:

- whether parents hold red books for children under 16;
- if they are willing to provide information from the red book as photographs or by data entry;
- how easy/costly it is to turn the photographs into useable data.

### 9.1 *Design*

Households, where the study team knew in advance there were children under 16 years, were randomised into two groups:

- Group 1 was sent a request in their pre-interview letter asking them to upload one page of each of their children under 16's red book. If they did not do this before the interview, they were asked again during their interview. The pre-interview letter was sent to all adults where we knew they had dependent children 4 weeks before the invitation to participate in the survey.
- Group 2 were not told of the request in advance but asked to upload photos from the red book during their interview.

Parents or Guardians with children under 1 were asked to upload the page/data from the measurement closest to when the child was 6-8 weeks, and for children over the age of 1 (but under 16) the measurements for the record closest to their first birthday was requested. During the photo upload, parents were asked for the child's name and date of birth, so the information could be attached to the correct child. Parents with multiple children were asked to upload each child's data separately. Parents were given £2 for each photo they uploaded.

In the household interview a responsible adult is identified for all children under 16. For ease, below we call the responsible adult a 'parent', although in some cases it may be a guardian. During the identified parent's interview, they were asked if they have a red book for each child. For those with a red book, if they were in group 1 they were asked if they had already uploaded a photo for all children. If not, or if they were in group 2, or the study team were not aware of children in the household before the interview, they were asked to upload the photographs as part of the interview. In the online version, the responsible adult does this themselves; in the face-to-face interview, the interviewer did it. The telephone protocol follows the web approach. Unfortunately, the questionnaire software employed by our fieldwork agency did not allow for files to be attached, and so this had to be scripted in separate software, to which parents or interviewers were ported if they agreed to upload photos.

If parents did not want to or could not upload a photo, but reported they had a red book, they were asked to record their children's weight and height at age 1 year. If the red book was not available, parents of children under 2 years, were asked to provide the



measurements and say how confident they were about them. Parents who didn't want to upload photographs were asked the reasons why.

A description of all the data extracted from the red book photographs and associated entered data can be found in chapter 19 (Benzeval and Payne, 2024, this volume).

## **9.2 Results**

In IP16 (University of Essex, Institute for Social and Economic Research, 2024), 365 people were identified as responsible adults for 600 children. Of them, 312 reported having a red book for at least some of their children (85%); 41 parents reported not having a red book, and 3 refused or didn't know. In total 147 adults said they would provide a photograph of red book data, 47% of those parents who reported having a red book for at least one child. Where measurement records were reported to be provided 146 parents provided photographs (70%), 48 entered information into the survey (23%); 2 parents did both, and 11 said they would provide data but did not.

This resulted in 209 data records, however, for 5 children both parents (from the pre-interview group) provided photographs, meaning there were measurement records for 204 unique children. Some parents provided photos for up to 4 children. Not all photographs had relevant data and not all parents who entered information provided relevant data. Weight data was provided for 156 children, 26% of the total children under 16 in interviewed households.

In terms of the experiment, 71 (27.2%) parents in interviewed households sent a pre-interview letter provided photographs of at least one of their children's red books and a further 10 (3.8%) entered data. In total 31% of parents in the pre-interview group provided red book or measurement information. Of the parents in group 2, whose households were not sent an advance letter, 62 (25.2%) provided photographs and a further 16 (6.5%) input measurements into the survey, i.e., a total of 31.7%. Overall, therefore, there was no meaningful difference in the provision of measurement information between those asked in advance and during the interview. Of those who did not use the red book, the main reason

for not doing so was not having access to it (65%) and secondly not trusting data security or feeling comfortable providing such data on their children (26%).

A wide range of red book formats were uploaded, presumably due to variations in format over time and between different health regions; health visitors recorded information by hand in very varied ways; and not all parents provided the pages requested. Given this it made it very hard to create a standardised data entry process, although a range of checks were included to verify the data entered where possible. A senior researcher therefore entered the data, so that judgements could be made about the information provided as it was entered. We concluded that, currently, it is not possible to automate the extraction of data from red books, and therefore it would not be possible to obtain measurement data at scale in this way.

As noted above parents with children under 1 were asked to upload the page/data from the measurement closest to when the child was 6-8 weeks, and for children over the age of 1 (but under 16) the measurements for the record closest to their first birthday was requested. Parents were provided with photographs of illustrative pages to help them identify the correct page. Of those providing photographs, 96% of parents uploaded a photograph that included the child's weight, although not necessarily for the requested age or nor from the requested page. The wrong page being uploaded, was more likely to be the case, where data collection was carried out by an interviewer, which suggests better training and guidance to interviewers may reduce the risk of this happening in future.

### **9.3 Conclusion**

Asking parents to provide measurement data from the red book or enter it into the survey manually, resulted in data for just over a quarter of children. There was no difference in the experimental treatment (i.e., whether the request was made before or during the interview). The workload associated with setting up the photograph upload process, and then entering the data from these photos, was significant. Data entry could not be automated, given the varied nature of red books and health visitors' entry of information. In

the long term it may be possible to link to the red book digitally, once all health regions have made the shift to e-records. Meanwhile, further experiments focused on encouraging parents to enter weight information using the red book may be the best way forward.

## **9.4    *References***

Benzeval, Michaela and John Payne. (2024). Child development measures from the ‘red book’. In Vine, J. et al. *Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data*, 148–155. Understanding Society Working Paper Series 2024–11. Colchester, Essex: Institute for Social and Economic Research, University of Essex.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## **10. The effect of incentives on the willingness to participate in a gamified app study**

*Jonathan Burton, Institute for Social and Economic Research, University of Essex*

*Annette Jäckle, Institute for Social and Economic Research, University of Essex*

*Mick Couper, University of Michigan*

### **10.1 Motivation**

As the use of apps becomes more prevalent, researchers are keen to use them to collect more, and different types of, data. The wide range of apps, and their functions, allows for a huge number of uses in collecting active or passive data. Whilst the data collected from these apps may be innovative, some of the standard questions of data collection are the same that are relevant to traditional methods, such as those around the selection and measurement effects of the data collection instruments.

Our interest is in the use of apps to collect additional data from respondents on a general population panel study. In past work, we have looked at different fieldwork protocols to encourage people to engage in app studies. In this chapter we are looking at the effect of incentives. In other research the level of incentive has not been particularly effective in persuading people to take part in app studies. In the Understanding Society Spending Study 1 we found that there was no difference in the proportion of respondents who downloaded an app between those who were offered £2 compared to £6 (Jäckle et al, 2019). In a study where we asked respondents to download an app and answer a short survey on well-being each evening for 30 days, offering a £10 bonus did increase the number of people who downloaded the app, but not on the number who used it (Jäckle et al, 2023). An incentive experiment in the IAB-SMART study also found no effect of incentives on app take-up (Haas et al, 2018; referenced in Kreuter et al, 2020). However, an experiment that offered €10 vs €20 to install a study app did lead to a slightly higher rate of installation for the larger incentive (Haas et al, 2021, Keusch et al, 2022). McCool et al. (2021) found that a conditional incentive of €20 for app use was more effective at encouraging app use than lower amounts

(€10, or €5 for registration + €5 for app use). In this experiment we test a greater difference in incentives: £10 vs £30, both conditional on using the app. We use data from an app study that was implemented in Wave 16 of the 2023 Understanding Society Innovation Panel survey (IP16) (University of Essex, Institute for Social and Economic Research, 2024). The IP is a probability sample of households in Great Britain.

## **10.2 *The Sea Hero Quest app***

During the IP16 interview, we asked respondents to locate, download, and use the Sea Hero Quest app. Sea Hero Quest is a multi-platform adventure game (available for Apple and Android phones and tablets) that was designed specifically for the Alzheimer's Research UK charity to help advance the understanding of spatial navigation, and therefore understand one of the first symptoms of dementia. Once logged in, the player was shown a map briefly, and then had to navigate a boat from the starting position to get to the flag that had been shown on the map (see screenshots in the Appendix, below). The app recorded how long it took for the player to get to the flag. The game is split into seventeen levels, with different environments, and gets more challenging at each level.

Further information about the Sea Hero Quest app study can be found in chapter 20 (Coutrot, Al Baghal and Spiers, 2024, this volume).

## **10.3 *Research questions***

- (1) Does participation depend in which mode of survey the respondent is invited to the app study?
- (2) Does the level of incentive affect the participation rate?
- (3) Does the level of incentive reduce non-participation bias?
- (4) Where in the process of installing and logging in to the app do we lose respondents?

## **10.4 Study design**

**Sample:** All respondents in the 2023 Innovation Panel survey who had completed at least one previous interview (n=2,694) were invited to install the Sea Hero Quest app. The Innovation Panel is part of Understanding Society: the UK Household Longitudinal Study. All members (aged 16 or over) of sample households are interviewed annually about their socio-economic situation including education, employment, income, health, housing, household membership and relationships. The design of the Innovation Panel is based on that of the main survey but fielded separately as a platform for methods testing and experimentation.

IP16 is a mixed-mode survey, and so those who do not complete in the mode to which they were issued are followed-up and invited to take part using a different mode. Web-first sample members who do not complete their interview online in the first five weeks are issued to CAPI interviewers to follow-up. Those who are issued to interviewers are invited to complete online during the reissue phase after 10 weeks. Towards the end of fieldwork, adults who had not yet responded were passed to interviewers to try and contact for a telephone (CATI) interview. Across all respondents, 75.4% completed their survey online, 21.6% completed in CAPI, and 3.0% completed by telephone (CATI).

**Experiments:** Our study involves two experiments. Treatment allocations were at the household level, such that all members of a household received the same treatment. The allocations for (2) were stratified by allocations to (1):

- (1) Survey mode of the annual interview. Around 60% of the sample were randomly allocated to be issued web-first, and 40% to CAPI-first. Non-respondents to the primary mode were followed up in the secondary mode plus CATI.
- (2) Respondent incentives: Half the sample were allocated to a conditional incentive of £10 and half to £30 if they downloaded and used the app. We defined using the app as logging in, completing the initial demographic questions, and completing the first level. The information about the app was not mentioned in any of the advance materials but was introduced for the first time during the individual adult interview.

Respondents were given their access code during the interview. If taking part on a mobile device, the respondent was given a link to the app in the appropriate app store. If they were taking part on a different device, or wanted to use the app on a different device, or were being interviewed by an interviewer, they were shown QR codes that would take them to the app in the Apple or Google app stores, or asked to search for the app in their app store.

## **10.5 Results**

### **RQ1: Does participation depend on the mode of the survey in which respondent are invited to the app study?**

We have found, in previous research, that there is often an effect of the mode of interview (self-completion online versus in-person face-to-face interviews) on the take-up of additional studies (e.g., app studies, giving biological samples) and giving consent to link administrative data to survey responses.

For the Sea Hero Quest app, we did not find this mode effect. Overall, 47.3% of those who took part at IP16 started using the app. For those who were issued CAPI-first, and who participated at IP16, 48.3% started using the app. For those issued Web-first, 46.8% started using the app ( $\text{Chi}^2 = 0.5317, p = 0.466$ ).

The difference in the level of app take-up was not statistically significant between those who completed their interview in CAPI (44.7%) or online (48.7%). However, both were statistically significantly higher than those who completed their IP16 interview in CATI (29.6%,  $\text{Chi}^2 = 13.4570, p = 0.001$ ).

### **RQ2: Does the level of incentive affect the participation rate?**

The incentive, which was paid conditional on playing the Sea Hero Quest app, had a significantly positive effect on take-up. Among those who were offered £10, 42.9% used the app. For those who were offered £30, this increased to 51.4% ( $p < 0.001$ ). Amongst those who started to use the app, the higher incentive was also more effective at encouraging

people to play through to the end of the game: 57.4% of app-users completed 100% of the game, compared to 47.4% of those who were offered £10 ( $p < 0.001$ ).

**Table 10-1 App take-up, and completion, by incentive group**

	N	% started	% completed if started	% completed of all respondents
£10	1,294	42.9	47.4	20.3
£30	1,400	51.4	57.4	29.5
Chi-square		19.3397	12.7110	31.8594
		$p < 0.001$	$p < 0.001$	$p < 0.001$

### **RQ3: Does the level of incentive reduce non-participation bias?**

Prior to introducing the Sea Hero Quest app to respondents, we asked whether they were gamers: “How often do you play games on any electronic device, for example a computer, console, tablet or mobile phone?” In past research we have found that the activities that people do on their smartphone may affect their willingness to participate in app studies. For example, those who already used apps for mobile banking were more likely to take part in a spending study (Jäckle et al, 2019). Since the Sea Hero Quest app is presented as a game, we thought that those respondents to the survey who did not play games may be less likely to take part.

We found that there was no statistically significant difference between regular gamers (play games at least several times a week) and occasional gamers (less frequently than several times a week) when it came to starting the app: 59.1% of regular gamers, and 54.5% of occasional gamers ( $\text{Chi}^2 = 2.8898$ ,  $p = 0.089$ ). However, both were significantly more likely to have started the app than non-gamers, but even in this group a third, 32.9%, started the game (both  $p < 0.001$ ).

For all levels of gamers, the higher incentive encouraged more people to start using the app, but it had the largest effect for occasional gamers. For non-gamers, the higher incentive of



£30 increased the proportion who used the app by 8.2 pp, and for regular gamers by 7.7 percentage points. For occasional gamers, the higher incentive increased the proportion by 14.6 pp. However, testing this interaction formally we find that the variation of the effect of the incentive with the level of gaming is not statistically significant. There was also no significant interaction of the incentive and respondent sex, age, and whether they used a smartphone for gaming.

The previous analysis looked at the effect of the incentive on the take-up of the app study. The higher incentive increases take-up, particularly for occasional users. Whilst this might be good, we should also consider the potential for the incentives to affect the bias in the app sample. Table 10-2, below, shows the sample composition for the full sample and the app users. For age and gaming experience, the app user sample over-represents younger (<60 years) respondents and those who are gamers and under-represents older respondents and non-gamers. Overall, for these characteristics, there is an average absolute bias of 5.7 percentage points. The higher incentive very slightly reduces the average absolute bias to 5.5, whereas the lower incentive is associated with a higher bias of 5.9.

**Table 10-2 Bias in the app sample compared to the full sample**

	Full sample: N	Full sample: %	Resp: %	Bias	Bias Z- stat	p-value
Female	1492	55.4	57.2	1.8	1.877	0.061
Age 17-30	412	15.3	18.6	3.3	5.580	<0.001
Age 31-40	405	15.0	20.0	5.0	9.703	<0.001
Age 41-50	393	14.6	18.4	3.9	7.004	<0.001
Age 51-60	499	18.5	20.3	1.8	2.499	0.012
Age 61-70	513	19.0	14.5	-4.5	-4.740	<0.001
Age 71-94	472	17.5	8.1	-9.4	-8.670	<0.001
Gaming: daily	675	25.1	31.8	6.7	10.072	<0.001
Gaming: several times a week	403	15.0	18.2	3.2	5.528	<0.001
Gaming: several times a month or less	488	18.1	20.9	2.8	4.065	<0.001
Gaming: never	1125	41.8	29.1	-12.7	-10.334	<0.001
Smartphone used for gaming: yes	975	36.2	49.2	13.0	20.143	<0.001
Average absolute bias				5.7		

#### **RQ4: Where in the process of installing and logging in to the app do we lose respondents?**

During the IP16 interview, we asked respondents about their progress in downloading the app. We combined these self-reported responses with information on whether we received data from the app, see Table 10-3 below. This was to identify where we were losing respondents. We find that just under 5 percent of respondents said that they didn't have a compatible device. We lost just under 30 percent because they were not willing to participate, with another 1.5 percent saying that they didn't try to locate the app. Just over 8 percent of respondents attempted to use the app, but could either not find it, or couldn't install it, or once installed could not log in to the app. Around 56 percent of respondents said that they had downloaded and logged in to the app, but for 8.5 percent we did not receive any data back, suggesting that they had not then used the app. We received data back from 47.3 percent of respondents. These results suggest that although we can make some advances in making it easier to locate, install, and log into the app, the biggest barrier is willingness to participate.

**Table 10-3 Drop-out during the app process**

<b>SHQ app outcome</b>	<b>Freq.</b>	<b>Percent</b>
No compatible device	129	4.8
Not willing to participate	798	29.6
Did not try	39	1.5
Did not find app	121	4.5
Could not install app	63	2.3
Could not log in	41	1.5
No app data received	229	8.5
App data received	1,274	47.3
Total	2,694	100.0

There are no significant differences between those who completed online or in CAPI, but those who completed in CATI were significantly more likely to say that they did not want to install the app. For those who did not want to download the app, the main reasons were that they did not want to participate in additional survey tasks, or they did not have time to take part, see Table 10-4 below.

**Table 10-4 Reasons for not wanting to download the app**

<b>Reasons</b>	<b>Freq</b>	<b>% All respondents</b>	<b>% Not want to d/l app</b>
I don't want to participate in additional survey tasks	260	9.7	28.5
Do not have time to take part	258	9.6	28.2
Not interested in answering additional questions on this topic	146	5.4	16
Do not want to take up storage space on my device	86	3.2	9.4
No smartphone or tablet that can download apps	75	2.8	8.2
Not able or confident to download apps onto my device	50	1.9	5.5
Not willing to share this kind of information	48	1.8	5.3
No internet access	33	1.2	3.6
Not confident that information would be held securely	28	1.0	3.1
<i>Total</i>		<i>2694</i>	<i>914</i>

## **10.6 Summary**

Unlike previous app studies, we did not find that the mode of issue had an effect on the rate of take-up. There was also no statistically significant difference by mode of interview for those who took part online and those who were interviewed face-to-face by an interviewer. Those who were interviewed by telephone were, however, less likely to take part in the app study. We also find that offering a large incentive (£30) rather than a more standard incentive (£10), increased app take-up. It also increased the rate at which those who used the app completed it. This effect was found for those who did not consider themselves gamers, as well as occasional and regular gamers. However, the interaction of the incentive and gaming experience, sex, and age indicated that the increased incentive was not

associated with significantly higher app use. The larger incentive slightly decreased the average absolute bias for the characteristics examined, but only by a small amount. The biggest barrier to increased participation is still the unwillingness to participate in the additional study.

## **10.7 References**

- Coutrot, A., Al Baghal, T., and Spiers, H. (2024). Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16. In Vine, J. et al. *Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data*, 156–164. Understanding Society Working Paper Series 2024–11. Colchester, Essex: Institute for Social and Economic Research, University of Essex.
- Haas, G. C., Kreuter, F., Keusch, F., Trappmann, M., and Bähr, S. (2018, August). What do researchers have to invest for collecting smartphone data. In Nonresponse Workshop August.
- Haas, G.-C., Kreuter, F., Keusch, F., Trappmann, M. and Bähr, S. (2021). Effects of incentives in smartphone data collection. In: Hill, C.A., Biemer, P.P., Buskirk, T.D., Japac, L., Kirchner, A., Kolenikov, S. & Lyberg, L.E. (Eds.) *Big data meets survey science. a collection of innovative methods*. Hoboken: Wiley, pp. 387–414.
- Jäckle, A., Burton, J., Couper, M.P., and Lessof, C. (2019) 'Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: coverage and participation rates and biases', *Survey Research Methods*, 13(1):23-44. <https://doi.org/10.18148/srm/2019.v1i1.7297>
- Jäckle, A., Burton, J., Couper, M.P., and Perelli-Harris, B. (2023) 'Participation of household panel members in daily burst measurement using a mobile app: effects of position of the invitation, bonus incentives, and number of daily questions', *Understanding Society Working Paper Series*, 2023-01. <https://www.understandingsociety.ac.uk/research/publications/547673>

Florian Keusch, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, Mark Trappmann, Stephanie Eckman, Non-Participation in Smartphone Data Collection Using Research Apps, *Journal of the Royal Statistical Society Series A: Statistics in Society*, Volume 185, Issue Supplement\_2, December 2022, Pages S225–S245, <https://doi.org/10.1111/rssa.12827>

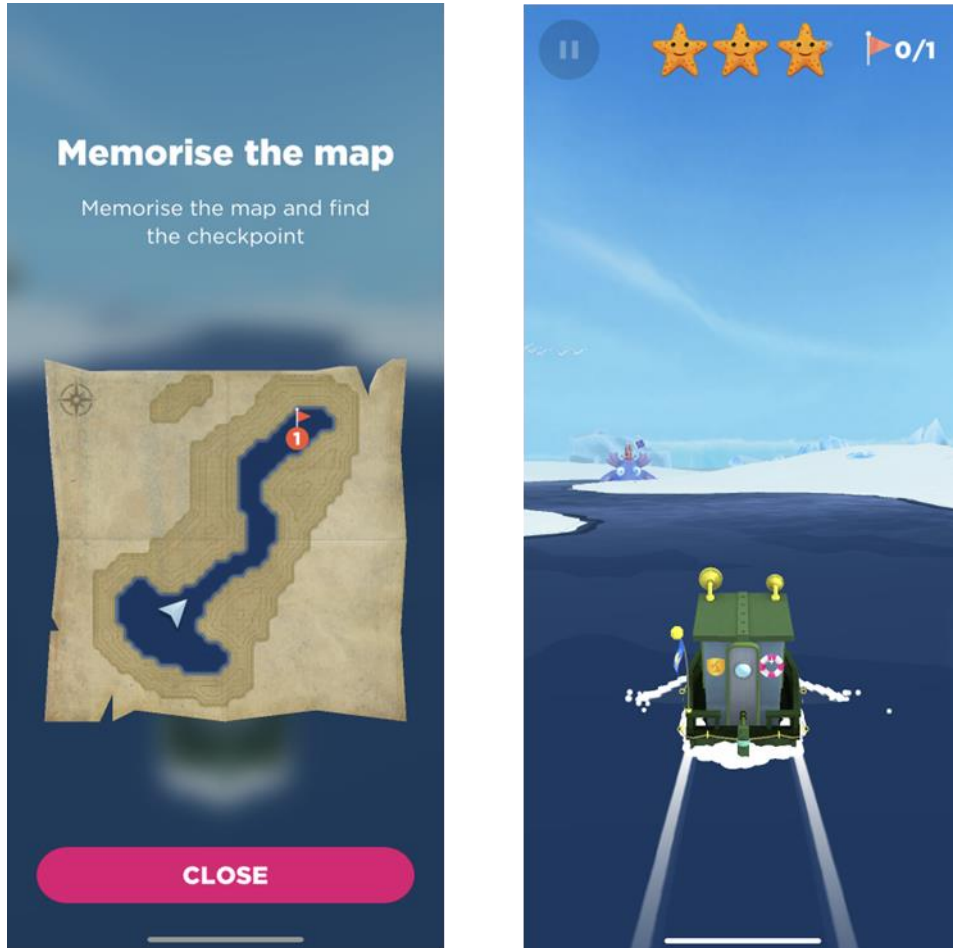
Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent. *Social Science Computer Review*, 38(5), 533-549. <https://doi.org/10.1177/0894439318816389>

McCool, D., Schouten, B., Mussmann, O., & Lugtig, P. (2021). An app-assisted travel survey in Official Statistics. Possibilities and challenges. *Journal of Official Statistics*, 37 (1): 149–170, <https://doi.org/10.2478/jos-2021-0007>.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## 10.8 Appendix

Figure 10-1 Screenshots from the app, showing the map and the game



## 11. Use of a government logo

*Jonathan Burton, Institute for Social and Economic Research, University of Essex*

### 11.1 Introduction

In Understanding Society, we send letters to adult sample members to ask them to take part in the study. For those who are issued web-first, the letter contains the URL for the survey, and the sample member's own log-in and password. If we have an email address for the sample member, we also send them an email with a direct link to their survey. For those who are issued CAPI-first, we let them know that an interviewer will soon be calling. This initial letter contains the incentive (gift-card) for previous-wave participants. Previous-wave non-responding adults are promised a gift-card if they complete their interview. The envelopes used for the letters use the Understanding Society name and logo, and the letters are addressed to the sample member. We are able to do this as a longitudinal study because we have interviewed at the household before and so we know who we are trying to contact.

This is not the case when we are trying to recruit a new sample. In the UK we use the Postcode Address File (PAF), a sampling frame of addresses, to draw a sample. This sampling frame does not contain the names of the people who live at the address, or any other contact details. This means that when trying to recruit a sample, we need to address the invitation letter to "The Resident". Unfortunately, this is how a lot of 'junk mail' is addressed, and so we suspect that a lot of the invitations to join the study end up in the bin, unopened.

In 2022 we started to recruit a new general population sample to boost Understanding Society. During the second year of the Wave 14 boost, we experimented with the use of a government logo on the invitation envelope. The aim was to encourage residents to open the envelope, so they could learn more about the study. This was carried out experimentally for one quarter (Q5 of Wave 14) for households issued January-March 2023. During this quarter, households who received the invitation letter in a standard Understanding Society envelope had a response rate of 20%. For households who had the letter in an envelope

that also included a government logo had a response rate of 28%. This was deemed a success and the government logos were used on the boost for the rest of the second year.

## **11.2 Methods**

The use of the government logo appeared to be successful in recruiting a new sample, for whom the name “Understanding Society” might not be significant, and where the recipient was just “The Resident”. We wanted to test whether it would have any effect on a more established sample – where the survey was known, and we had the names of the sample members. We decided to experiment with the use of the government logo at IP16 (University of Essex, Institute for Social and Economic Research, 2024). The original Innovation Panel was established in 2008. Since that wave, there have been a number of refreshment samples, to maintain the sample size. Refreshment samples were recruited at IP4 (2011), IP7 (2014), IP10 (2017), IP11 (2018), and IP14 (2021).<sup>4</sup> There was no refreshment sample at IP16, and so the issued sample comprised only individuals that had already taken part in the study and where we had a name. Many of the sample members had been part of the study for a long time, although around a quarter were part of the IP14 refreshment sample. Our expectation was that the presence of a government logo on the envelope would not encourage greater participation, since the logo was intended to persuade people to open the envelope. The personal addressed envelope and the study logo, we believed, would be sufficient for that task.

## **11.3 Results**

Table 11-1, below, shows the adult response rates for the treatment group (government logo) and the control group (no government logo). The response for those who had the

---

<sup>4</sup> The ‘early’ refreshment sample at IP11 was to increase the sample size in advance of IP12, which was used to test different procedures for a biomarker collection wave.



standard Understanding Society envelope was 57%. This was not statistically different to those who had the envelope that included the government logo (55%). Those adult sample members who were in the £30 incentive group were slightly more likely ( $p=0.09$ ) to take part if they were in the control group (i.e., no government logo) than if they were in the treatment group (73% compared to 67%). There was no difference between adults who had taken part in the previous year, compared to those who were non-respondents. Looking at the sample 'cohort', those who had been recruited in the most recent refreshment sample – at IP14 – were significantly more likely to take part if they were in the control group ( $p=0.004$ ). Those who joined as part of the IP10 refreshment sample were marginally more likely to take part if they were in the government logo group ( $p=0.09$ ).

**Table 11-1 Adult response rates for treatment and control groups**

	<b>Government logo</b>	<b>No Government logo</b>	<b>n</b>
Response	1,390 55%	1,437 57%	2,827 56%
<b>Incentive</b>			
£20	1,131 53%	1,134 54%	2,265 53%
£30*	259 67%	303 73%	562 70%
<b>Previous-wave outcome</b>			
Respondent	1,104 82%	1,163 83%	2,267 82%
Non-respondent	267 25%	247 24%	514 24%
<b>Sample origin</b>			
IP1 original	361 61%	383 63%	744 62%
IP4 refreshment	184 66%	164 62%	348 64%
IP7 refreshment	176 56%	188 58%	364 57%
IP10* refreshment	116 57%	118 49%	234 53%
IP11 refreshment	177 45%	164 44%	341 44%
IP14*** refreshment	376 51%	420 58%	796 55%

\* p<0.1; \*\*\* p<.01

## **11.4 Conclusions**

Overall, the government logo did not have a significant effect on response to the study. There are suggestions that for the most recently recruited refreshment sample, it may have harmed response. The decision has been to not use government logos on the existing

Understanding Society sample. However, we are planning boost samples in the coming years, and we expect to use government logos in the initial recruitment wave.

## **11.5 *References***

University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, [DOI: https://doi.org/10.5255/UKDA-SN-6849-16](https://doi.org/10.5255/UKDA-SN-6849-16).

## **12. Establishing a databank of smallest effect sizes of interest (SESOI) for prominent psychological constructs**

*Karen L. Mansfield, Oxford Internet Institute, University of Oxford*

*Andrew K. Przybylski, Oxford Internet Institute, University of Oxford*

### **12.1 Introduction**

This study addressed a fundamental problem in psychological research, which is how to identify when changes in subjective wellbeing, mood, or affect are meaningful. For example, a university might offer free yoga sessions for several weeks, measuring student and staff wellbeing before and after the intervention. A proportion of those who attended the yoga sessions might record higher scores at the end of the intervention, but how much higher do the scores need to be at post-test compared to pre-test for this to be considered a meaningful effect? Scores at post-test might be considered significantly higher statistically speaking, but with large sample sizes even tiny effect sizes can exceed common thresholds. To what extent might a change in scores be of practical significance as well as statistical significance? A popular approach to this problem in psychology is based on the standardised effect size defined by Cohen (1988), calculated by dividing the average change by the standard deviation, with the resulting standardized effects interpreted as small (0.2), medium (0.5), or large (0.8). However, even Cohen himself expressed that these cut-offs were arbitrary, only to be used when no better indicators were available.

The current study uses an alternative approach, proposed by Anvari and Lakens (2021), who used the example of estimating the smallest effect size of interest for the positive and negative affect scales (PANAS, Watson et al., 1988). It draws from prior work used in clinical research, aiming to define a Minimal Important Difference (MID), by comparing the change in outcome scores to a subjective anchor (King, 2011). In clinical research, the anchor might be a clinician's judgement of whether or not a patient has improved, but for clinical or non-clinical studies, participants' own subjective experience (e.g., worse, the same, better) can also be used as an anchor. Considering the subjective nature of psychological constructs like

wellbeing or life satisfaction, using participants' own experience of change is logically appealing, and it highlights how what constitutes a meaningful effect should vary between psychological constructs and how they are measured.

Building on the Anvari and Lakens example, the current study aimed to assess the smallest effect size of interest for changes in the measure of overall life satisfaction implemented annually in the Understanding Society surveys.

## **12.2 Methods**

The Understanding Society Innovation Panel (University of Essex, Institute for Social and Economic Research, 2024) presented the opportunity to analyse data on life satisfaction from adults and adolescents in two consecutive years (T1=2022; T2=2023), along with key demographics, and to add a bespoke anchor measure at T2.

*Measures:* The measure for which this study estimates the smallest effect size of interest is "satisfaction with life overall", which is measured using a 7-point Likert scale, for which responses range from "completely dissatisfied" to "completely satisfied", with "Neither satisfied nor dissatisfied" at the centre. The anchor item then asked participants: "Compared to how satisfied you were **with your life overall** ... [when we last interviewed you], how would you rate your satisfaction with your life overall **these days**?". Responses for the anchor item were scored on a 5-point Likert scale, with options: "1: Much lower", "2: Slightly lower", "3: About the same", "4: Slightly higher", and "5: Much higher", such that the *slightly lower/higher* groups could be used to estimate the smallest noticeable changes in life satisfaction.

*Raw and standardised effect sizes:* To calculate the smallest effect size of interest using the anchor method by Anvari and Lakens (2021), we first computed a difference score in life satisfaction for each participant (life satisfaction at T2 – life satisfaction at T1). Positive values therefore indicate that life satisfaction has increased over time. Next, we calculated the average difference from T1-T2 for each of the 5 anchor response categories. As well as the raw average differences, we calculated standardized effect sizes (Cohen's *d*, with and

without accounting for correlations), and effect sizes expressed as the percentage of maximum possible (POMP) score units, along with their 95% confidence intervals. For further details on the value and application of POMP effect sizes, please refer to Anvari and Lakens (2021) and Cohen et al. (1999). To calculate the smallest subjectively experienced difference in POMP score units, we first converted the scores for T1 and T2 into POMP units using the formula:  $((\text{observed score} - \text{min score}) / (\text{max score} - \text{min score})) * 100$ ; where min score is the minimum possible score on the life satisfaction scale (1); and max score is the maximum possible score (7). The smallest subjectively experienced difference was then calculated using the POMP units in the same way as for the raw score units. Next, we reverse coded the change scores for those who reported their life satisfaction to be “much lower” or “slightly lower”, and used a paired t-test to confirm whether changes for the “slightly lower” and “slightly higher” groups were symmetrical, to determine whether it is appropriate to merge these groups and calculate an overall effect size for meaningful changes in either direction.

*Quality considerations:* We ran several tests to assess the reliability of the anchor item as an estimate, also following recommendations by Anvari and Lakens (2021). First, we used paired t-tests to confirm that the change scores for the ‘slightly lower’ and ‘slightly higher’ groups, separately and merged, differed significantly from the ‘about the same’ group. Next, we tested for correlation between anchor item scores and change scores, as well as correlation between the anchor item and life satisfaction at T2, and between the anchor item and life satisfaction at T1. Ideally, the anchor item should correlate more strongly with change scores (T2-T1) than with scores at T2 or T1 (Cella et al., 2002; Devji et al., 2020).

### **12.3 Results**

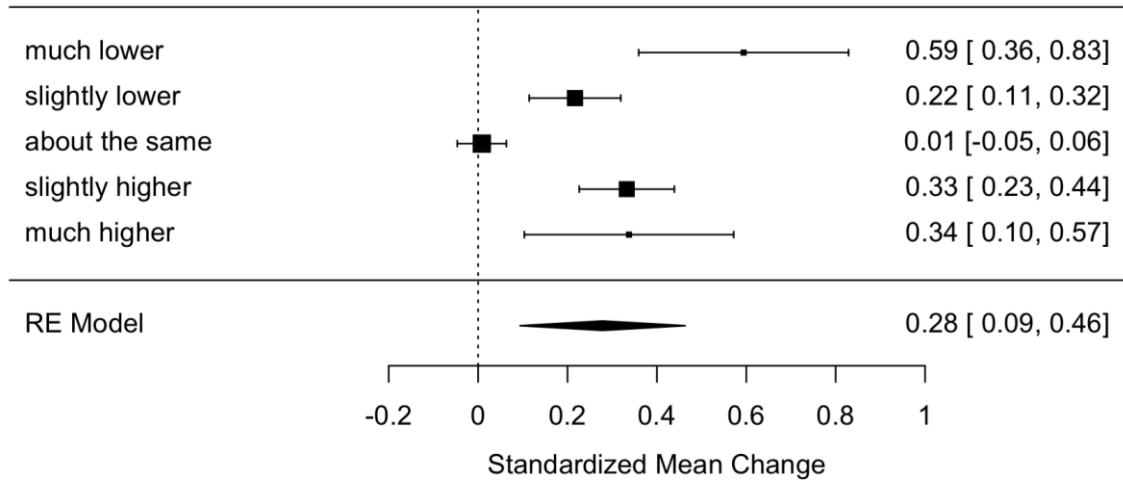
*Sample:* After merging data from 2022 and 2023, excluding participants for whom there was no life satisfaction data at either T1 or T2, and excluding participants who either did not answer the anchor question or answered, “Don’t know”, a total of 2164 participants were included in complete case analysis. Participants’ ages at T2 ranged from 17-94 years (mean 54 years), of which 976 (45%) were men.

**Table 12-1 Summary statistics according to response on the anchor item, with raw (mean differences), standardized ( $d_z$  and  $d_{av}$ ), and POMP effect sizes, with 95% confidence intervals.**

Anchor	N	T1: M(SD)	T2: M(SD)	Mean difference	Cohen's $d_z$	Cohen's $d_{av}$	POMP
Much lower	82	3.51(1.68)	2.56(1.47)	-0.95[-1.30;-0.60]	-0.6[-0.83;-0.36]	-0.6[-0.84;-0.36]	-15.85[-0.84;-0.36]
Slightly lower	375	4.52(1.42)	4.24(1.31)	-0.28[-0.41;-0.15]	-0.22[-0.32;-0.11]	-0.20[-0.30;-0.11]	-4.67[-0.30;-0.11]
About the same	1275	5.38(1.26)	5.39(1.24)	0.01[-0.05;0.07]	0.01[-0.05;0.06]	0.01[-0.04;0.06]	0.16[-0.04;0.06]
Slightly higher	358	5.39(1.28)	5.77(0.90)	0.38[0.26;0.50]	0.33[0.23;0.44]	0.34[0.23;0.45]	6.38[0.23;0.45]
Much higher	74	5.42(1.41)	5.91(1.20)	0.49[0.16;0.82]	0.34[0.11;0.57]	0.37[0.11;0.63]	8.11[0.11;0.63]

*Smallest effect sizes of interest:* Raw, standardised, and POMP effect sizes with confidence intervals are presented in Table 12-1. Important to note is that participants who said they felt ‘about the same’ demonstrated no overall difference between their life satisfaction scores at T1 and T2, which implies that we can interpret the effect sizes in Table 12-1 for the ‘slightly lower’ and ‘slightly higher’ groups as relevant to meaningful changes in each direction, and don’t need to calculate instead their difference from the ‘same’ group (Redelmeier et al., 1993, 1997). Similar to the finding using the PANAS (Anvari & Lakens, 2021), the raw effect size appeared smaller for participants who reported their life satisfaction to be slightly lower than for participants who reported their life satisfaction to be slightly higher. However, after reversing the change scores for the groups who reported their life satisfaction to be lower (see Figure 12-1), a paired t-test comparing the change scores for the ‘slightly lower’ and ‘slightly higher’ groups confirmed that the changes were symmetrical ( $t(728)=1.14, p=.25$ ). As such, it is appropriate to merge the two ‘slightly’ changed groups to calculate an overall effect size for changes in either direction, which results in an overall raw effect size of 0.33 ( $CI_{95\%}[0.24; 0.42]$ ,  $SD = 1.22$ ) and an overall standardised effect size (Cohen's  $d_z$ ) of 0.31 ( $CI_{95\%}[0.24; 0.38]$ ).

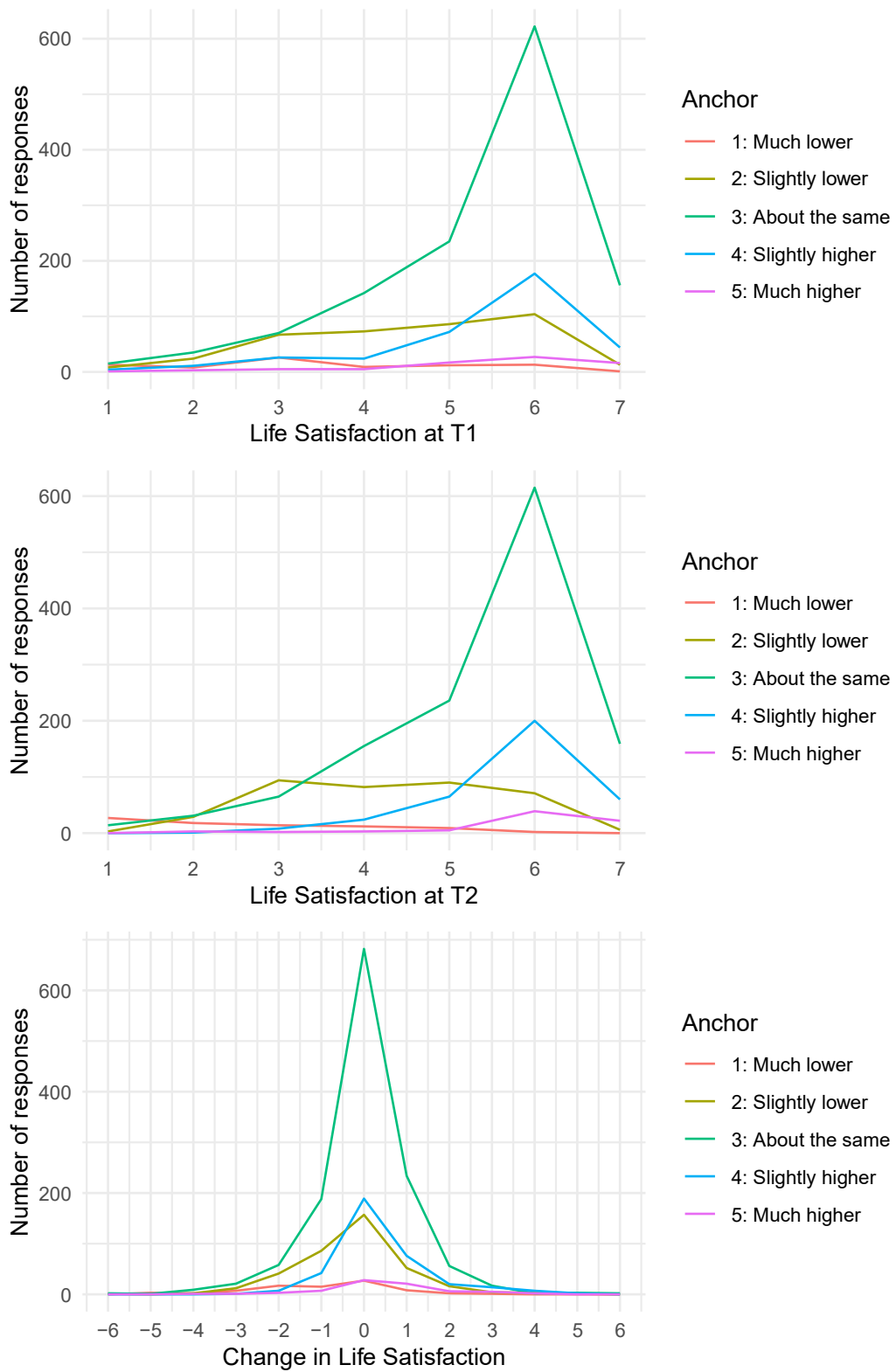
**Figure 12-1. Forest plot of standardised effect sizes, following reverse coding of life satisfaction change scores for those who responded “much lower” or “slightly lower”, including result of a random effects meta-analysis.**



*Quality considerations:* Paired t-tests confirmed that the change scores for those who reported their life satisfaction to be “about the same” differed significantly from those who reported it to be “slightly lower” ( $t(555)=3.67, p<.001$ ) or “slightly higher” ( $t(568)=5.45, p<.001$ ), also when combining the slightly changed groups ( $t(1434)=5.81, p<.001$ ). When testing correlations, anchor item scores were indeed positively correlated with change scores ( $r = .22, CI_{95\%} [.18, .26]$ ), but the correlation between anchor item scores and life satisfaction at T2 ( $r = .46, CI_{95\%} [.42, .49]$ ) was stronger ( $r_{dif} = -.23, CI_{95\%} [-.27, -.19]$ ; Zou, 2007). The anchor item was also correlated with life satisfaction at T1 ( $r = .26, CI_{95\%} [.22, .30]$ ), but the anchor item’s correlation with life satisfaction at T2 was stronger than its correlation with life satisfaction at T1 ( $r_{dif} = .19, CI_{95\%} [.16, .23]$ ; Zou, 2007). Exploratory analyses to assess distributions of responses in more detail (Figure 12-2) revealed that the largest group of participants (>600) reported being “mostly satisfied” with life overall, both at T1 and at T2. Similarly, for the largest group (>600), life satisfaction scores did not change from T1 to T2.



**Figure 12-2 Distributions of life satisfaction scores at T1 and T2, and change scores, by response to the anchor item.**



## 12.4 Discussion and conclusions

In this study we applied the method proposed by Anvari and Lakens (2021) by adding a subjective anchor item to the Understanding Society Innovation Panel survey in 2023 (University of Essex, Institute for Social and Economic Research, 2024) and used participants' responses to estimate a smallest effect size of interest for meaningful changes in the measure of overall life satisfaction over a period of one year. Based on the symmetry of change scores for subjectively perceived 'slight' changes in either direction, and on their significant difference from change scores for participants who reported their life satisfaction to be 'about the same', we estimated an overall, raw, smallest effect size of interest of 0.33 ( $CI_{95\%}[0.24; 0.42]$ ,  $SD = 1.22$ ), and an overall standardised effect size of 0.31 ( $CI_{95\%}[0.24; 0.38]$ ).

Notably, responses to the anchor item seemed to reflect life satisfaction at T2 more strongly than changes in life satisfaction from T1 to T2, at least in this particular setting. Advice reported by Anvari and Lakens (2021) suggests that the correlation between the anchor score and change scores should be around 0.5, and this should be stronger than the correlation between the anchor item and life satisfaction at T2 or T1, for the estimated effect size to be considered reliable (Cella et al., 2002; Devji et al., 2020). While Anvari and Lakens were able to meet these criteria in two studies assessing short-term changes in positive and negative affect, the current study highlighted potential limitations to the approach.

One account for the anchor item primarily reflecting life satisfaction at T2 is that the length of time between annual surveys makes it more difficult for participants to remember exactly how they felt last time they were interviewed. Having just answered the life satisfaction question, participants might use that as a reference and guess how they responded one year earlier. In contrast, the strong correlation between the anchor item and changes in PANAS scores found by Anvari and Lakens (2021) likely reflects their short 2–5-day delay between T1 and T2. Another account is that because our analysis used a single item measure of current life satisfaction, rather than a multi-item scale like PANAS, makes the

scores at T1 and T2 more prone to noise, and the change scores especially noisy. Indeed, Anvari and Lakens noted that correlations with their anchor item were lower for individual PANAS items than for the aggregate scores. More research can be done to test the extent to which the reliability of a subjective anchor depends on the length of time between T1 and T2, as well as on the number of items in the specific measure, and later to assess the extent to which the value of a subjective anchor varies according to other factors such as the characteristics of the sample.

Keeping this limitation in mind, our study offers a new way for researchers using Understanding Society to evaluate changes in life satisfaction between surveys and their association with other factors, besides statistical significance and crude advice on effect sizes (Cohen, 1988). Our analysis provides a smallest subjectively noticeable difference – a standardised effect size of 0.31 ( $CI_{95\%}[0.24; 0.38]$ ) – for researchers to consider when evaluating the practical significance of differences in life satisfaction outcomes in the Understanding Society data and comparable surveys.

## **12.5 References**

- Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology*, *96*, 104159. <https://doi.org/10.1016/j.jesp.2021.104159>
- Cella, D., Hahn, E. A., & Dineen, K. (2002). Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Quality of Life Research*, *11*(3), 207-221. [https://doi.org/10.1016/s0149-2918\(02\)85118-9](https://doi.org/10.1016/s0149-2918(02)85118-9)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315-346. [https://doi.org/10.1207/S15327906MBR3403\\_2](https://doi.org/10.1207/S15327906MBR3403_2)

- Devji, T., Carrasco-Labra, A., Qasim, A., Phillips, M., Johnston, B. C., Devasenapathy, N., ... & Urquhart, O. (2020). Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*, 369. <https://doi.org/10.1136/bmj.m1714>
- King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics & Outcomes Research*, 11(2), 171– 184. <https://doi.org/10.1586/erp.11.9>
- Redelmeier, D. A. (1993). Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Archives of Internal Medicine*, 153(11), 1337–1342. <https://doi.org/10.1001/archinte.153.11.1337>
- Redelmeier, D A, Bayoumi, A. M., Goldstein, R. S., & Guyatt, G. H. (1997). Interpreting small differences in functional status: The Six Minute Walk test in chronic lung disease patients. *American Journal of Respiratory and Critical Care Medicine*, 155(4), 1278-1282. <https://doi.org/10.1164/ajrccm.155.4.9105067>
- University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>
- Watson, D., Anna, L., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, 12(4), 399–413. <https://doi.org/10/fmb3nm>

## 13. Anticipated gender discrimination in the labour market

*Gurleen Popli, Department of Economics, University of Sheffield*

*Anita Ratcliffe, Department of Economics, University of Sheffield*

### 13.1 Summary and key findings

Theory and evidence suggest that anticipating labour market discrimination influences decision-making, including educational and occupational choices. However, there is limited data available on the prevalence and correlates of these expectations. To assess the prevalence of anticipated gender discrimination in the labour market, we proposed a question to be included in the Innovation Panel of the UK Household Longitudinal Study (UKHLS), asking the respondents if they think that their gender will make it harder for them to get a job. The question is administered to respondents aged 16 - 59 in the adult survey, and to respondents of all ages (11 to 15) in the youth survey. Our **key findings** suggest:

- While women are most likely to anticipate gender discrimination in the labour market, a non-trivial proportion of men anticipate gender discrimination.
- For women, anticipated gender discrimination correlates with demographic characteristics; women with young children are more likely to report anticipated discrimination compared to men, and ethnic minority women are twice as likely to report anticipated discrimination relative to white women.
- For men, anticipated gender discrimination correlates with job characteristics; with men working part-time reporting much higher levels of anticipated discrimination relative to men working full-time, and men in female-dominated industries are least likely to report anticipated discrimination.

### 13.2 Motivation

The last few decades have seen an increase in female participation in both higher education and the labour market, yet the gender wage gap persists despite improvements in

education and labour market experience. While there are various explanations for the gender wage gap, one explanation is gender discrimination in the labour market, and our research seeks to understand whether women expect to face such discrimination. This matters because scholars across various disciplines suggest a link between anticipated labour market discrimination and the educational investments of minority groups (Arcidiacono et al. 2010; Lang and Manove 2011; Dickerson et al. 2024). It is also likely that anticipated labour market discrimination influences job search activity by altering expected payoffs.

To date, there is very little empirical evidence, especially from observational data, on the prevalence and consequences of anticipated gender discrimination in the labour market, owing to a lack of data.<sup>5</sup> Next Steps, a survey of adolescents in England, is one of the few surveys which asks respondents about anticipated ethnic discrimination in the labour market.<sup>6</sup> Dickerson et al. (2024) use the anticipated discrimination question in Next Steps to show that ethnic minority students anticipating labour market discrimination have higher educational achievement at age 16.

To assess the prevalence of anticipated gender discrimination in the labour market, we proposed that the same question that appears in Next Steps - albeit specific to gender - be included in the Innovation Panel of UKHLS (University of Essex, Institute for Social and Economic Research 2024). This question allows us to look at the prevalence and correlates of anticipated discrimination for adolescents and adults, specific to the labour market.

---

<sup>5</sup> There is some experimental evidence available on the impact of anticipated discrimination on the choices made by individuals, see Charness et al. (2020).

<sup>6</sup> Next Steps (formerly Longitudinal Survey of Young People in England) is a large national survey of over 15,000 children born between 1<sup>st</sup> September 1989 and 31<sup>st</sup> August 1990. Adolescents were first interviewed in 2004 when they were on an average 14 years old, and then annually till 2010, with the final interview in 2015 when the respondents were on an average 25 years old. An anticipated discrimination question was asked of the respondents in 2005, when the respondents were on an average 15 years old.

### **13.3 Question included in the Innovation Panel (IP)**

If still in full-time education:

‘Looking ahead to after you leave education, do you think that your gender will make it harder for you to get a job?’

Responses: Yes / No / Don't know / Inapplicable - not planning to work

If completed full-time education (and regardless of current labour market status):

‘Looking ahead to the next time you might apply for a job, do you think that your gender will make it harder for you to get a job?’

Responses: Yes / No / Don't know / Inapplicable - not planning to work

The question is administered to respondents aged 16 - 59 in the adult survey, and to respondents of all ages (11 to 15) in the youth survey. The wording is slightly different for respondents still in full-time education (i.e., all respondents in the youth sample, and some respondents in the adult sample) to reflect the fact that respondents are starting from different positions when looking ahead. However, the main thrust of the question is the same, as are the possible answers, and we treat this as a single question.

### **13.4 Sample and preliminary analysis of responses**

There are 2880 respondents in the adult IP, with 1788 aged 16-59, and 131 respondents in the youth IP totalling 1919 relevant respondents. One respondent in the adult survey is aged 60 but answered the question, we exclude this response.

Table 13-1 presents responses in the adult and youth sample, as well as combined responses. Non-response rates (i.e., ‘refused’) are extremely low (always less than 1%) though there is a higher proportion of ‘inapplicable’ responses, with a higher proportion of

these in the youth sample.<sup>7</sup> Approximately 7% of the adult sample responded ‘not planning to work’ with no such responses in the youth sample. Notably, the vast majority of respondents do not anticipate gender discrimination in the labour market but approximately 8% of respondents do anticipate gender discrimination, with a slightly higher proportion of ‘yes’ responses in the youth sample. A small proportion of respondents (approximately 2%) reply with ‘don’t know’, with this figure largely driven by respondents in the youth sample.

**Table 13-1 Response to the anticipated discrimination question**

<b>Response</b>	<b>Adult (16-59)</b>	<b>Youth (11-15)</b>	<b>Total</b>
Not planning to work	122 (7%)	0 (0%)	122 (6%)
No	1465 (82%)	87 (66%)	1552 (81%)
Don’t Know	23 (1%)	23 (18%)	46 (2%)
Yes	138 (8%)	14 (11%)	152 (8%)
Refused	13 (1%)	1 (1%)	14 (1%)
Inapplicable	27 (2%)	6 (5%)	33 (2%)
<i>Observations</i>	1788	131	1919

*Note:* percentages provided in parentheses

Table 13-2 provides response rates by gender. Women are most likely to indicate they anticipate gender discrimination in the labour market (11%) but a non-negligible proportion of men (4%) also anticipate gender discrimination. Girls in the youth sample are most likely to indicate they anticipate labour market discrimination (20%). Respondents in the adult sample rarely respond with ‘don’t know’ but this response is just as common as ‘yes’ responses for girls in the youth sample and is also a common response for boys in the youth sample (approximately 14%).

---

<sup>7</sup> In the adult sample a high proportion (14 of 27) of those who responded ‘inapplicable’ are currently in paid employment, it is possible that these individuals do not perceive themselves as looking for a job in the near future.



**Table 13-2 Response to the anticipated discrimination question, by gender**

Response	Adults		Youth		Total	
	Males	Females	Males	Females	Males	Females
Not planning to work	49 (6%)	73 (7%)	0 (0%)	0 (0%)	49 (6%)	73 (7%)
No	679 (87%)	786 (78%)	53 (80%)	34 (52%)	732 (87%)	820 (76%)
Don't Know	6 (1%)	17 (2%)	9 (14%)	14 (22%)	15 (2%)	31 (3%)
Yes	33 (4%)	105 (10%)	1 (2%)	13 (20%)	34 (4%)	118 (11%)
Refused	5 (1%)	8 (1%)	0 (0%)	1 (2%)	5 (1%)	9 (1%)
Inapplicable	8 (1%)	19 (2%)	3 (5%)	3 (5%)	11 (1%)	22 (2%)
<i>Observations</i>	780	1,008	66	65	846	1073

Note: percentages provided in parentheses

### **13.5 Analysis**

In the analysis that follows, we consider how anticipated gender discrimination varies with demographic and job-related characteristics. To simplify this analysis, we drop 'not planning to work', 'inapplicable' and 'refused' responses, which reduces the sample to 1750 individuals (1626 in the adult sample and 124 in the youth sample). We also merge 'yes' and 'don't know' responses to create a binary variable indicating whether an individual has entertained the possibility of facing gender discrimination in the labour market. It should be borne in mind, however, that 'don't know' responses represent a larger share of responses in the youth sample. According to this definition, 11% of respondents report that they anticipate discrimination due to their gender (15% of women versus 6% of men).

### 13.6 Demographic characteristics

We consider how anticipated gender discrimination varies with demographic characteristics, such as age and ethnicity, as this is potentially informative of intersectionality between gender and other protected characteristics.

Figure 13-1 below shows the percentage of male and female sample reporting yes to the question, by age group. Younger women are more likely to anticipate gender discrimination, with this proportion falling until age 30, and somewhat levelling out afterwards. As we cannot distinguish between age and cohort effects, we do not know if this is because younger cohorts have greater awareness of gender issues compared to older cohorts, or if experience in the labour market as respondents' age provides a different perspective on this. A similar pattern is observed for men, but they are also more likely to anticipate gender discrimination in later life, which narrows the gap between men and women.

**Figure 13-1 Anticipates gender discrimination by gender and age group**

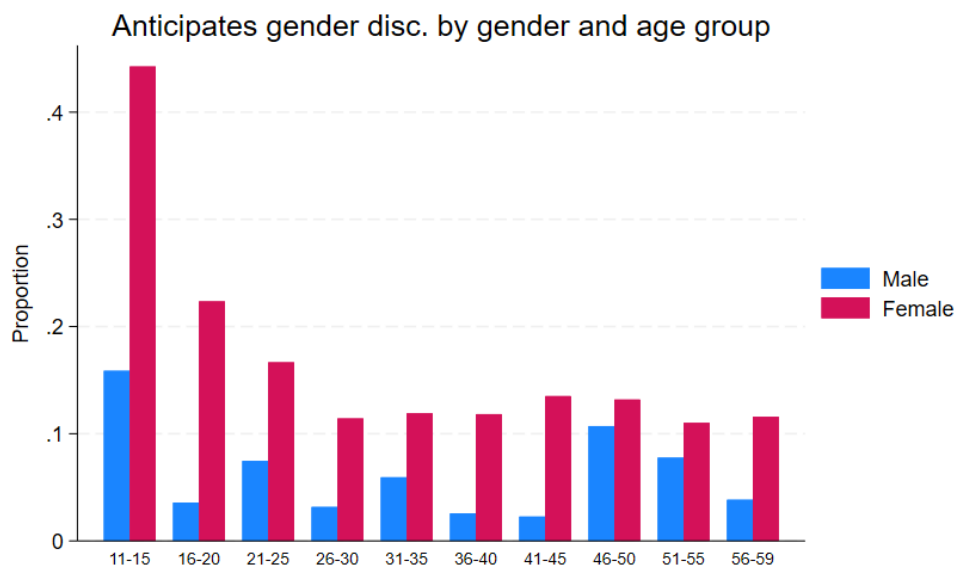
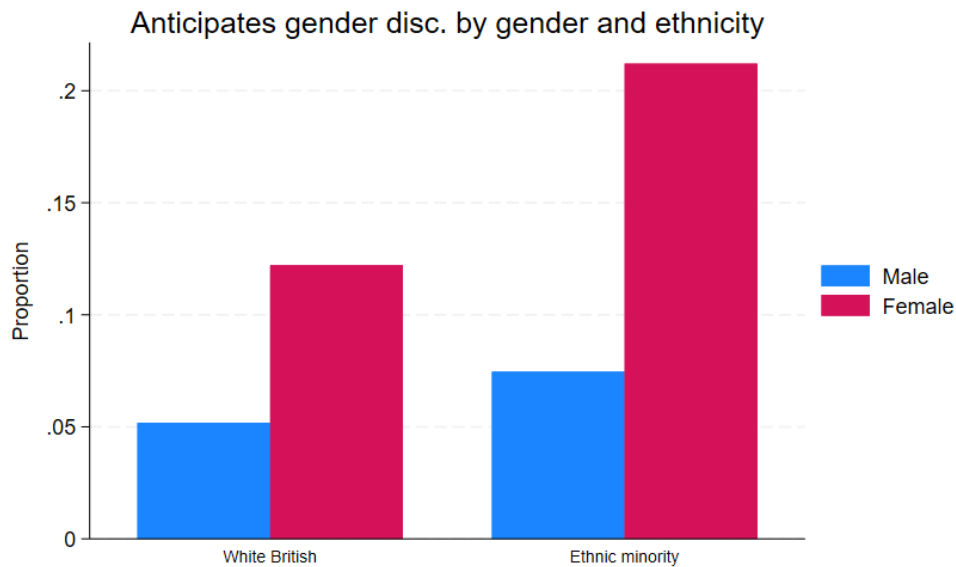


Figure 13-2 shows how anticipated gender discrimination in the labour market varies by ethnicity, based on a sample of 986 respondents for whom we have data on ethnicity in the adult sample, of whom 17% have an ethnic minority background (we do not disaggregate further due to small sample sizes). Ethnic minority women are almost twice as likely to

anticipate gender discrimination compared to white women. Ethnicity data is not available in the youth sample.

**Figure 13-2 Anticipates gender discrimination by gender and ethnicity**

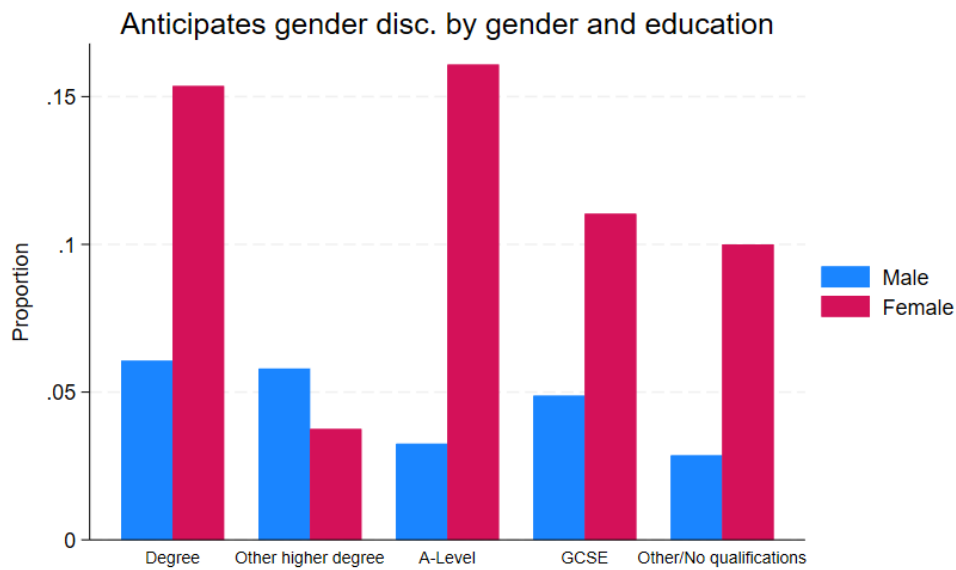


We also consider if anticipated gender discrimination varies by highest qualification level (adult sample only) where we have the required information for 1236 respondents, see Figure 13-3. Education may affect knowledge of labour market discrimination and/or labour market experience, both of which may influence whether people anticipate gender discrimination. Broadly, women with higher qualifications are more likely to anticipate gender discrimination in the labour market, though this is not uniformly the case, as women with an 'Other higher degree' are least likely to anticipate gender discrimination.<sup>8</sup>

---

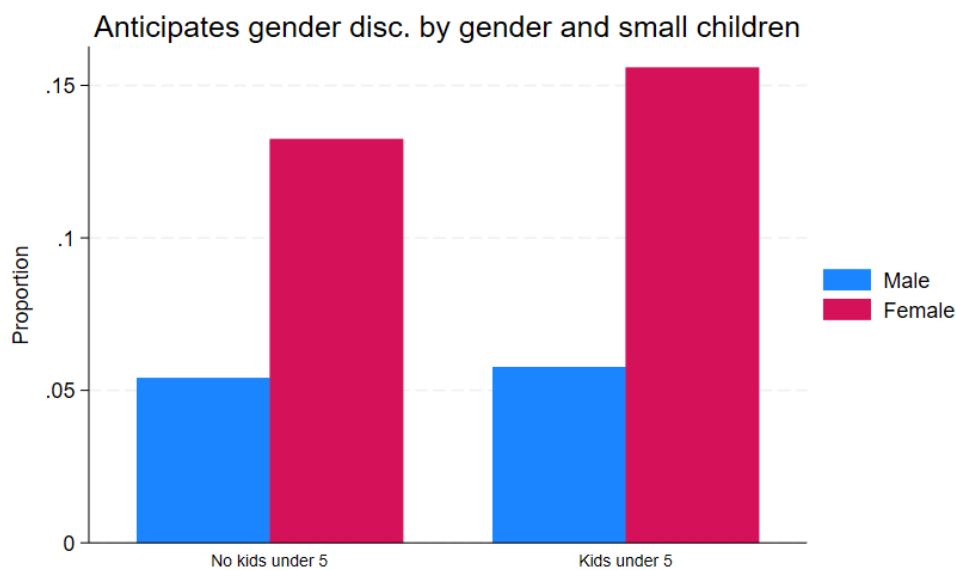
<sup>8</sup> 'Other higher degree' includes all those who have diplomas in higher education, a nursing or other medical qualification, and those with teaching qualifications. Over 40% of the respondents with 'other higher degree' work in the education or health sector, majority (78%) of which are women.

**Figure 13-3 Anticipates gender discrimination by gender and education**



In Figure 13-4 we consider how anticipated gender discrimination in the labour market varies according to whether respondents have children aged under 5 (based on adult sample). Women with small children are slightly more likely to anticipate gender discrimination compared to women without (despite those aged 16-20 being least likely to have small children and most likely to anticipate gender discrimination), with very similar proportions for men regardless of whether they have small children.

**Figure 13-4 Anticipates gender discrimination by gender and small children**

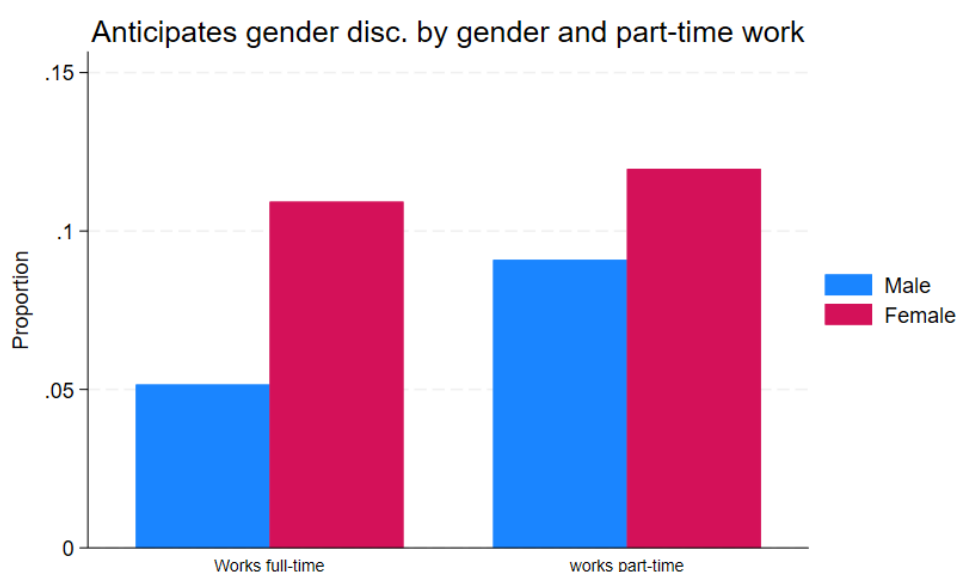


### 13.7 Job characteristics

Finally, we consider whether anticipated gender discrimination in the labour market correlates with job characteristics, based on the adult sample whose main economic activity is in employment, self-employment, on maternity-leave or in work-related training (totalling 1317 individuals).

We start by considering part-time work status given its link to parenting status. We define part-time work as working less than 30 hours per week for the sample of people report job hours (1159 individuals). In this sample, 67% of women and 8% of men work part-time, see Figure 13-5. Interestingly, working part-time makes men more likely to anticipate gender discrimination in the labour market, with very little difference for women.

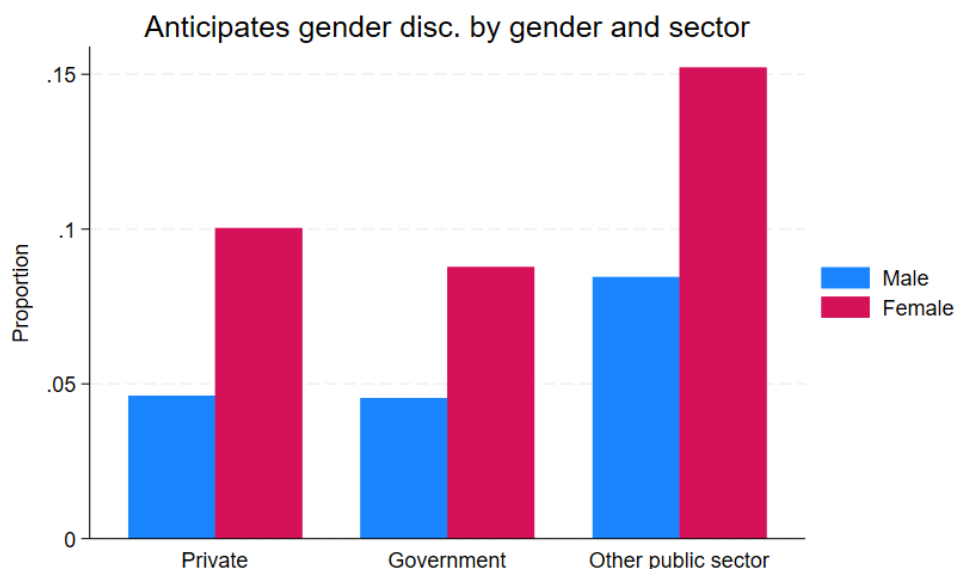
**Figure 13-5 Anticipates gender discrimination by gender and part time work**



We next consider how anticipated gender discrimination in the labour market varies by sector of employment where this information is available for 1002 individuals. We compare private sector versus other sectors where we categorise individuals as working for the government (i.e. central/local government) and other public sector organisations (i.e. health authority/NHS, charities and other kind of organisation). Results are reported in Figure 13-6. Interestingly, both men and women working in other public sector roles are most likely to

anticipate gender discrimination while women working for the government are slightly less likely to indicate they anticipate gender discrimination compared to women working in the private sector.

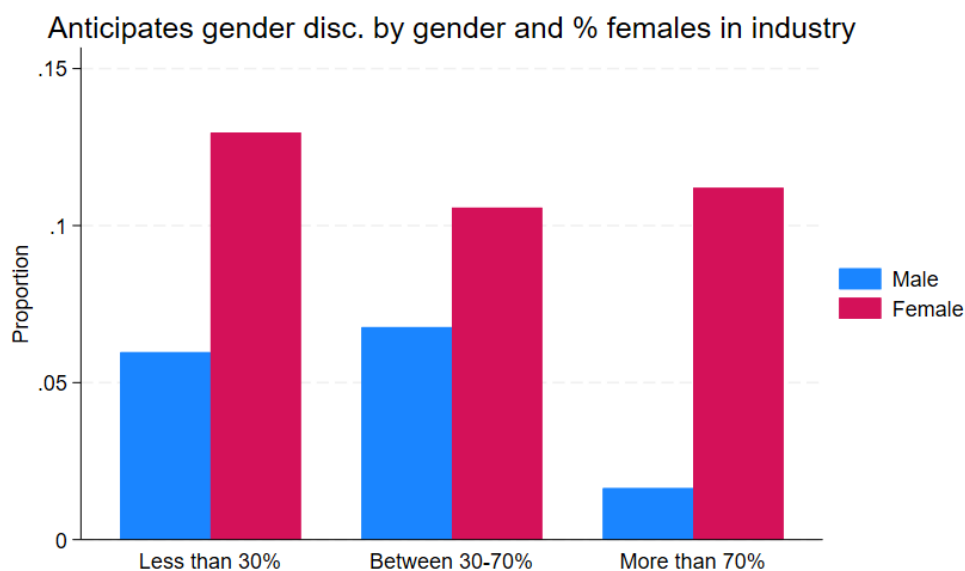
**Figure 13-6 Anticipates gender discrimination by gender and sector**



Finally, we consider whether people are more likely to anticipate gender discrimination when working in male/female dominated industries. Industry of employment is collected from Waves O and N of the Innovation Panel and therefore may not reflect current industry of employment for some individuals. In total, industry of employment is available for 983 individuals. We calculate the % of females working in an industry based on Nomis workforce jobs by industry (SIC 2007) and sex at March 2023. Education and Human health and social work activities are classified as having more than 70% female workers. Mining and quarrying; Manufacturing; Electricity, gas, steam and air conditional supply; Water supply, sewerage, waste management and remediation activities; Construction, and Transportation and Storage are classified as having less than 30% female workforce. Women are most likely to anticipate gender discrimination in the labour market when working in a male-dominated industry and least likely to anticipate gender discrimination when working in a mixed-gender industry, see Figure 13-7. Men are least likely to anticipate gender discrimination when

working in a female-dominated industry and most likely to anticipate gender discrimination when working in a mixed-gender industry.

**Figure 13-7 Anticipates gender discrimination by gender and proportion of women employed in industry**



### **13.8 Discussion and conclusion**

Based on the responses from the adult and the youth sample in the IP survey we find that while women are more likely to anticipate gender discrimination in the labour market, a non-trivial proportion of men also anticipate gender-based discrimination in the labour market. In the adult sample, where detailed demographic and job characteristics are available, we find that women are more likely to report anticipated discrimination based on demographic characteristics, whereas men are more likely to anticipate gender-based discrimination based on job characteristics.

Initial analysis of the IP data on anticipated question is interesting but also raises lots of questions, such as – what drives anticipated gender discrimination for young women?; and how much of this is driven by actual experiences of gender discrimination? For example, ethnic minority women are twice as likely to report anticipated discrimination relative to white women. This could be a reflection of the current experience, where evidence suggest

significant 'ethnic penalty' in labour market outcomes, especially for women (Zwysen et al. 2021; Khattab and Hussein 2018). Similarly, women with small children (below the age of 5) are more likely to anticipate gender discrimination, this again is consistent with the evidence on the 'motherhood penalty' faced by women (Kleven et al. 2019).

Findings with respect to job characteristics are also interesting. Men working part-time report higher levels of anticipated gender discrimination, relative to men working full-time. While there exists extensive research on women in part-time jobs, evidence on the composition of male part time workers and the implications of this for men is scarce and less well understood; see Warren (2022) for a recent discussion on part time work by gender. The public sector, often viewed as a more equitable employer, is associated with higher reports of anticipated discrimination by both men and women. Industry also warrants further investigation. Women consistently reported higher expectations of discrimination across all industries. Interestingly, men in female-dominated industries reported much lower levels of anticipated discrimination. These findings raise important questions for future research on anticipated gender discrimination and job characteristics. The impact of part-time work, sector (public vs. private vs. government), and industry (gender composition) deserve closer examination.

In this report we present an initial analysis of the prevalence and correlates of anticipated gender discrimination. The findings conform with the existing evidence on observed gender-based discrimination in the labour market, and open up new avenues which need further investigation. The next steps in this research would be to understand the coping mechanisms adopted by individuals in response to this anticipated gender discrimination.

### **13.9 References**

Arcidiacono, P., Bayer, P. & Hizmo, A. (2010). Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics* 2(4), 76–104.



- Charness, G., Cobo-Reyes, R., Meraglia, S., & Sánchez, Á. (2020). Anticipated discrimination, choices, and performance: Experimental evidence. *European Economic Review*, 127, 103473.
- Dickerson, A., Ratcliffe, A., Rohenkohl, B., & Van de Sijpe, N. (2024). Anticipated labour market discrimination and educational achievement. *Journal of Economic Behavior & Organization*, 222, 375-393.
- Khattab, N., & Hussein, S. (2018). Can religious affiliation explain the disadvantage of Muslim women in the British labour market?. *Work, Employment and Society*, 32(6), 1011-1028.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, 109, 122-126.
- Lang, K. & Manove, M. (2011). Education and labor market discrimination. *American Economic Review*, 101(4), 1467–96.
- University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>
- Warren, T. (2022). Work-time, male-breadwinning and the division of domestic labour: male part-time and full-time workers in unsettled times. *Sociology*, 56(1), 72-96.
- Zwysen, W., Di Stasio, V., & Heath, A. (2021). Ethnic penalties and hiring discrimination: Comparing results from observational studies with field experiments in the UK. *Sociology*, 55(2), 263-282.

## 14. Domestic workers and platforms

*Marion Lieutaud, London School of Economics and Political Science*

### 14.1 Introduction and context

The aim of this Innovation Panel questionnaire module was to identify casual and fragmented uses of paid housework by households, something poorly captured by Understanding Society and survey data in general. This new data allows us to better understand important aspects of inequality in modern modes of social reproduction (who can outsource at least part of the housework? How widespread has this become in the rise of the gig economy?).

It is likely that many households have a regular but light (less than 10-20h a month) use of paid housework. This type of small 'gig' domestic work is increasingly common and disproportionately performed by migrant and ethnic minority women (Farris 2020). It is likely that it is also more and more often mediated through online platforms (see Ticona and Mateescu (2018); Muldoon (2021)). Digital labour platforms for domestic and care work have seen tremendous growth in recent years (ILO 2021), but understanding exactly how and how many households and workers use them, as well as what kind of working conditions they foster, has proven very difficult in the existing data landscape. Fragmented and 'gig' domestic work is the object of rising academic and political attention. But thus far, it is poorly captured by survey data in general, and by Understanding Society in particular, as one can only know that a task is outsourced to paid help if it fully outsourced (e.g.: who does the cooking? - paid help *only*). This leaves the regular but not systematic use of paid housework invisible. Families who, for instance, have a cleaner come in every second week or every month for a deep clean, will (legitimately) answer that most of the cleaning is performed by household members. By contrast, the new Innovation Panel module questions can also capture this fragmented or gig-type paid domestic work.

This is important in a context where the outsourcing of domestic work has been presented as a tool for gender equality in the household, freeing middle-class and upper-class women

from the duty to perform reproductive labour unpaid. But this data also has important implications for understanding how inequalities in social reproductive labour are being re-shaped between those - often underpaid and precarious (Cox 2006) - who perform that labour and those who can afford (albeit occasionally) to outsource it (Bhattacharya 2017). This new data can help us better understand the prevalence of different kinds of recruitment, including online platforms.

## **14.2 *Methods and data***

The Innovation Panel module was entitled 'Regular uses of paid domestic work - considering platforms' (University of Essex, Institute for Social and Economic Research 2024). It includes the following set of questions: Number of people doing paid housework for the household. How many hours of paid housework a week in total. Finally, they were asked which platform or app that was, if applicable.

The first questions have been designed to enhance comparability with other international surveys - especially the Gender and Generation Survey questionnaire. The focus on means of recruitment, platforms and work hours are new additions.

## **14.3 *Analysis***

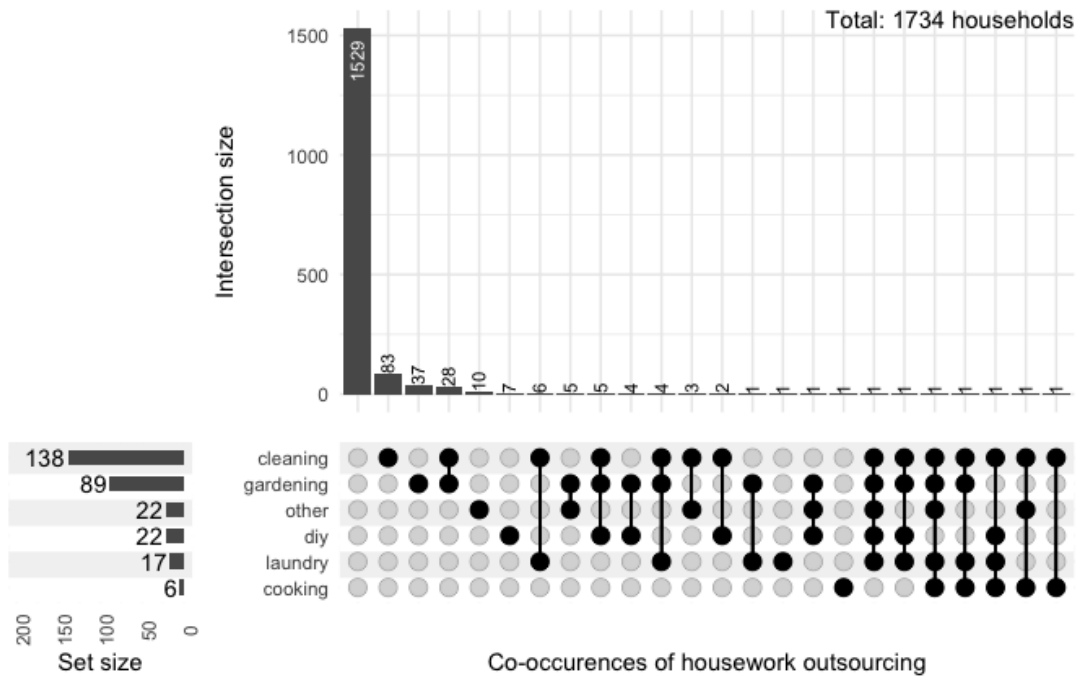
### **Who pays for (some) housework?**

The vast majority of households - 88% - do not outsource any housework at all in a typical month. Among those who regularly pay for some housework help (n=205), the most commonly outsourced housework is cleaning/tidying (8% of all households (138 households) in the dataset). The second most common is gardening. Whereas having a cleaner or a gardener is not altogether rare, paying for a cook appears very selective. Figure 1 shows the distribution and co-occurrence of different types of housework outsourcing, with the highest bar representing the most common scenario, i.e. households that do not outsource any task. The second most common scenario is households outsourcing only cleaning/tidying or

only gardening. As we move towards the right of the barplot, we see increasingly rare configurations where households outsource multiple types of housework, e.g. cleaning + gardening or cleaning + laundry + gardening, etc...

**Figure 14-1 Co-occurrences of housework outsourcing**

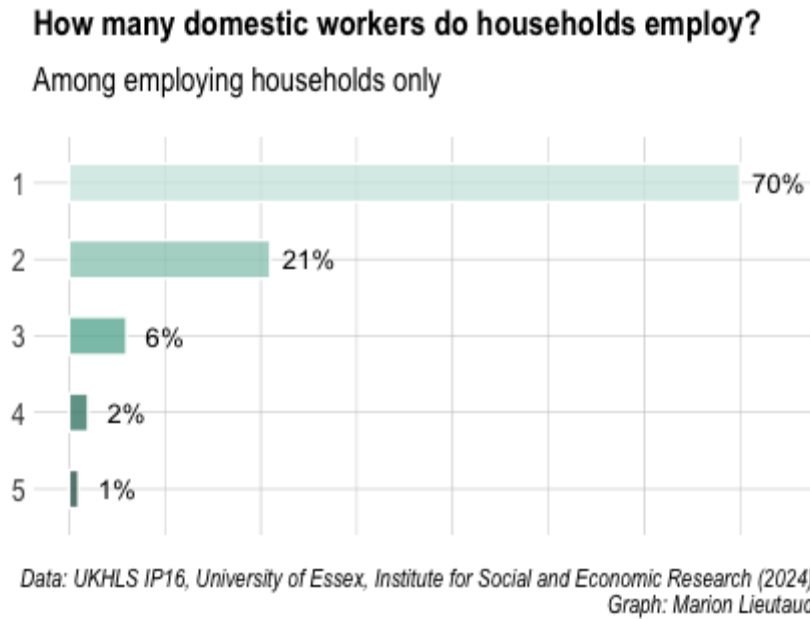
### What housework do British households pay for?



Data: UKHLS IP16, University of Essex, Institute for Social and Economic Research (2024) | Graph: Marion Lieutaud

## Number of employed domestic workers per household

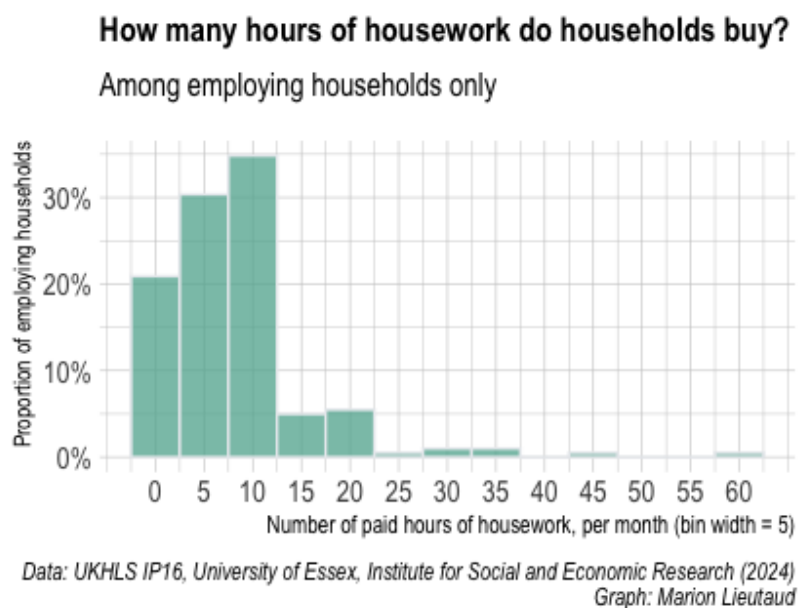
Figure 14-2 Numbers of domestic workers per household



Most employing households hire only one domestic worker - typically a cleaner. Around 30% of employing households pay more than one worker.

## How much housework do people pay for?

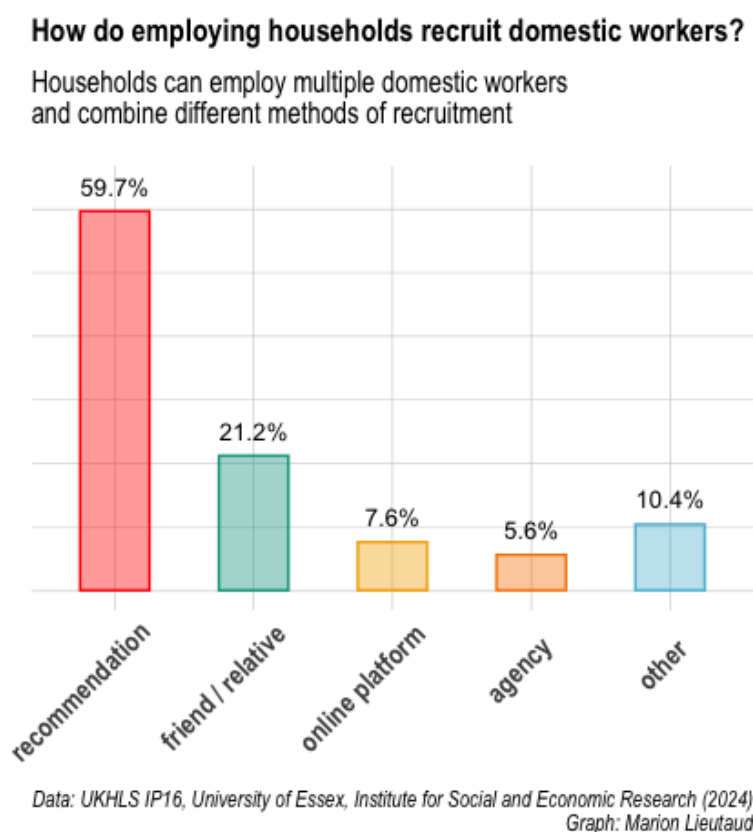
Figure 14-3 Distribution of number of hours of paid domestic work per employing household



In a typical month, 20% of employing households pay for 1-4 hours of housework. 30% outsource 5-9 hours. 35% outsource 10-14h. 5% pay for 15-19h, and another 5% for 20 to 24 hours per month, so roughly 5 hours per week. Very few households outsource more than 24 hours per month. The median among employing households is 6 hours per month, and the mean is 8 hours.

## Methods of recruitment

**Figure 14-4 Methods of recruitment of paid domestic workers used by employing households**



The most common method of recruitment of paid domestic worker is through recommendation by someone known to the household. 60% of domestic workers employed by households in the IP16 sample were found this way. Paid housework is also often performed by friends, relatives or acquaintances - further blurring the formal/informal boundaries around domestic and care work - with over 20% of workers falling into this

category. Agencies likely play a bigger role in mediating high-end domestic work (full-time, in-house, especially for super-rich globalised households, see e.g. Delpierre (2022)), but among employing households overall, agencies mediated the recruitment of only 6% of workers. Online platforms/apps and websites account for 8% of recruitment (n=22). Informal and direct methods of recruitment (friends and recommendations) thus continue to dominate paid domestic work, although digital labour platforms now represent a substantial share of the work and workers.

#### **14.4 References**

- Bhattacharya, Tithi. 2017. *Social Reproduction Theory. Remapping Class, Recentring Oppression*. London: Pluto Press.
- Cox, Rosie. 2006. *The Servant Problem: Domestic Employment in a Global Economy*. London: I.B. Tauris, I. B. Tauris.
- Delpierre, Alizée. 2022. *Servir Les Riches*. Paris: La Découverte.
- Farris, Sara R. 2020. "The Business of Care: Private Placement Agencies and Female Migrant Workers in London." *Gender, Work, and Organization* 27 (6): 1450–67.
- ILO. 2021. "Making Decent Work a Reality for Domestic Workers: Progress and Prospects Ten Years After the Adoption of the Domestic Workers Convention, 2011 (No. 189)." Report. International Labour Organisation.  
[http://www.ilo.org/global/topics/domestic-workers/publications/WCMS\\_802551/lang--en/index.htm](http://www.ilo.org/global/topics/domestic-workers/publications/WCMS_802551/lang--en/index.htm).
- Muldoon, James. 2021. *Platform Socialism: How to Reclaim Our Digital Future from Big Tech*. London: Pluto Press.
- Ticona, Julia, and Alexandra Mateescu. 2018. "Trusted Strangers: Carework Platforms' Cultural Entrepreneurship in the on-Demand Economy:" *New Media & Society*.

University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, [DOI: https://doi.org/10.5255/UKDA-SN-6849-16](https://doi.org/10.5255/UKDA-SN-6849-16).



## 15. Numeracy and the long-term future

*Lieke Voorintholt, University of Groningen*

*Adriaan R. Soetevent, University of Groningen*

*Gerard J. van den Berg, University of Groningen, University Medical Center Groningen, and IFAU Uppsala*

### 15.1 Introduction

Many policy decisions have an enormous impact on the welfare of future generations, for instance, by addressing intergenerational challenges such as climate change, nuclear waste storage, and technological risks. At the same time, many people find it difficult to think about the future beyond their lifespan (Tonn, Hemrick, & Conrad 2006), which complicates overseeing intergenerational challenges. Still, these policy decisions are driven by people's support for them. This makes it relevant to investigate people's grasp of the long-term future, especially given the limited attention this topic has received in the literature.

For the near future, several studies have already shown that the perception of time and growth influences behaviour and valuation. In a study that relates numerical misperception to the exponential growth pattern of the coronavirus, Lammers, Crusius, & Gast (2021) show that correcting people's bias increases support for social distancing. Moreover, on a personal consumption level, people's vision of future events is more strongly explained by their subjective perception of time rather than their actual time preferences (Bradford, Dolan, & Galizzi 2019).

Using newly designed question modules in the Understanding Society Innovation Panel (University of Essex, Institute for Social and Economic Research 2024), we investigate people's conceptual grasp of the long-term relative to the short-term future by comparing their performance on numeracy questions across three different timespans. Next, we determine the effect of framing the numeracy questions in a specific context such as climate change. As such, we contribute to existing research on framing effects in both the time

preference and the climate change literature (e.g. Johannesson & Johansson 1997; Kip Viscusi & Zeckhauser 2006).

Our questions are inspired by the related literatures on financial literacy (e.g., Lusardi & Mitchell 2008, Kaiser & Lusardi 2024) and exponential-growth bias (EGB) (e.g., Stango & Zinman 2009). The construct 'basic financial literacy' is included in several household panels and consists of five multiple-choice questions, the first involving interest compounding. This concept is a well-established predictor of various (financial) behaviours (Van Rooij et al. 2011, Kaiser & Lusardi 2024). Similarly, the EGB literature studies people's understanding of the compounding of interest and its relation to economic outcomes (Wagenaar & Sagaria 1975, Stango & Zinman 2009, Levy & Tasoff 2016).

Unlike EGB research, which typically employs open-ended questions asking respondents to estimate savings after a certain number of 'periods' (of unspecified length) with a given interest rate (Levy & Tasoff 2016), we use multiple-choice questions. This is in line with the financial literacy literature and reduces the burden on respondents. To the original financial literacy question's timespan of 5 years, we add variants for 30 and 100 years to measure whether participants find it more difficult to apply the same principle of interest compounding to time periods further in the future.

By exploring people's conceptual understanding of long-term processes, we take a different approach than most studies on intergenerational valuation, which usually use the concept of time preferences. Unlike time preferences for personal consumption, such intergenerational time preferences are difficult to elicit. Since outcomes (partly) lie beyond people's lifespan, the measurement of subjective valuation of such outcomes relies on stated preferences or hypothetical scenarios (Cropper, Aydede, & Portney 1994; Frederick 2003; Voorintholt 2020).

## **15.2 Methods**

Our new modules each consist of a one-sentence introduction followed by three multiple-choice sub-questions. The principle of interest compounding forms the basis for all our

questions. The financial numeracy module frames its questions around savings growth using a 5-year timespan to introduce the concept, with extensions for 30 and 100 years. Similarly, the environmental numeracy module, designed to be mathematically equivalent to the financial module, tests the understanding of interest compounding applied to sea level rise due to climate change instead of savings growth. Answer options are designed to capture respondents' understanding of exponential growth and allow us to identify whether a mistake is due to overestimation or underestimation. Following standard practice in financial literacy research, we can also simply categorize responses as correct or incorrect.

We employ a mixed research design that combines within- and between-individual comparisons to compare performance across modules. Twenty percent of the sample encounters both module types for within-subject comparisons, and the remaining 80% are evenly split, with half only exposed to the financial numeracy module and the other half to the environmental numeracy module, facilitating between-subject comparisons. Each respondent will face their module(s) either early or later in the self-completion part of the questionnaire and this placement is randomized to control for order effects.

Below, we list all questions for both the financial (a) and environmental (b) module types, with correct answers italicized for reference.

(a) Imagine a savings account with a current balance of **£100**. The interest rate is fixed at **1% per year**. After **5 years**, how much money will be in the account if no money is withdrawn?

(b) Imagine a certain island that currently lies **100 cm** below sea level. As a result of climate change the sea level rises by **1% per year**. After **5 years**, how many centimetres below sea level will the island be?

1. 101 or less
2. *More than 101, but no more than 106 (correct answer  $\approx 105.1$ )*
3. More than 106, but no more than 111
4. More than 111

(a) After **30 years**, how much money will be in the account if no money is withdrawn?

(b) After **30 years**, how many centimetres below sea level will the island be?

1. 110 or less
2. More than 110, but no more than 130
3. *More than 130, but no more than 150 (correct answer  $\approx$  135)*
4. More than 150

(a) After **100 years**, how much money will be in the account if no money is withdrawn?

(b) After **100 years**, how many centimetres below sea level will the island be?

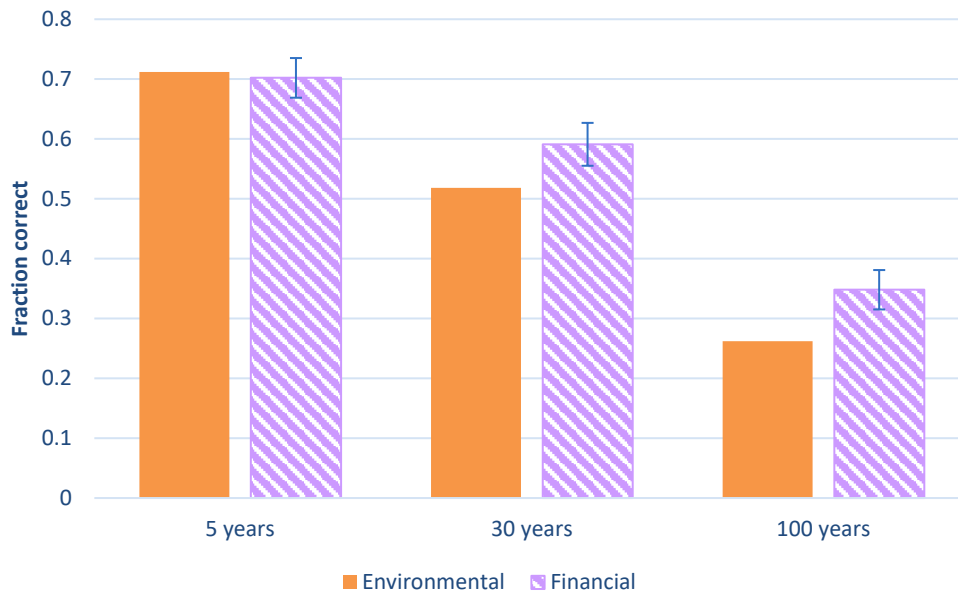
1. 150 or less
2. More than 150, but no more than 225
3. *More than 225, but no more than 300 (correct answer  $\approx$  270)*
4. More than 300

### **15.3 Results**

In this section, we present basic statistics of respondents' numeracy. Starting with Figure 15-1, we display a bar chart that shows the share of respondents who correctly answered specific numeracy questions, represented as fractions on the Y-axis. The bars are categorized by question timespan (5, 30, and 100 years) and module type (environmental or financial), shown on the X-axis.

The chart reveals a clear trend: respondent performance decreases as the timespan of the question increases, which reflects that respondents have more difficulty applying interest compounding to longer periods. While the majority of respondents still correctly answer the 30-year version, there is a noticeable drop in the fraction who correctly answer the 100-year version. Finally, it seems that the financial variant of the numeracy question improves performance for the 30- and 100-year versions.

**Figure 15-1 Fraction of correct answers per question**



Note: This figure is based on all 2943 respondents that are also included in Table 15-1 and Table 15-2. The Financial bars include a 95% confidence interval for the effect of financial framing as compared to environmental framing for a specific timespan.

Next, we pool respondents' performance across all three timespan versions (5, 30, and 100 years) to analyse overall trends. We summarize the fraction of questions correctly answered and, separately, the fraction of questions where respondents overestimated or underestimated the correct answer. We again split these statistics by environmental or financial framing and present them in Table 15-1.

Consistent with the trends observed in Figure 15-1, the average fraction of correctly answered questions is notably lower for the environmental framing compared to the financial framing. Interestingly, the difference is not equally driven by the fraction of questions overestimated as underestimated; respondents in the environmental module are more likely to underestimate rather than overestimate the correct answers. This pattern seems to indicate a context-related bias on performance, which we will explore in future work on this project.

**Table 15-1 Average performance per question module**

	Environmental Mean (SD)	Financial Mean (SD)	Env. – Fin. Diff. ( <i>p</i> -value)
Fraction correct	0.50 (0.31)	0.55 (0.32)	-0.05*** (<0.001)
Fraction overestimated	0.16 (0.29)	0.21 (0.32)	-0.05*** (<0.001)
Fraction underestimated	0.34 (0.32)	0.25 (0.29)	0.10*** (<0.001)
Observations	1482	1461	2943

Note: Summary statistics are reported only for respondents who answered all three questions per module.

This includes both the respondents who faced just one out of two modules (between-design group) and the respondents who faced both modules (within-design group). \*\*\*  $p < 0.001$

Table 15-2 presents the estimates from a simple regression of the fraction of correctly answered numeracy questions on module type (environmental vs. financial) and module placement (early vs. late in the questionnaire). Our results are separated by our two different estimation samples, the addition of controls, and the use of individual fixed effects for the within-design sample.

The results indicate that the environmental numeracy module results in a lower fraction of correctly answered questions than the financial numeracy module. This is the case both for respondents randomly assigned to one of the two modules (columns 1 and 2), and for respondents facing both modules (columns 3 and 4), although the effect appears stronger for the former group. The estimates for our early placement indicator suggest that facing a module earlier rather than later in the questionnaire reduces the fraction of correctly answered questions. The consistency of this effect across different models implies that it is not solely attributable to the set-up of our within-design where respondents face the same calculations in two modules.

**Table 15-2 Effects of numeracy question framing and placement**

	(1)	(2)	(3)	(4)
	Fraction of correctly answered questions			
Environmental	-0.06*** (0.01)	-0.06*** (0.01)	-0.03 (0.01)	-0.03* (0.01)
Early module	-0.03*	-0.03*	-0.04**	-0.04**

Constant	(0.01) 0.56***	(0.01) 0.60***	(0.01) 0.65***	(0.01) 0.58***
	(0.01)	(0.03)	(0.05)	(0.01)
Controls	No	Yes	Yes	No
FE	No	No	No	Yes
Observations	1961	1961	982	982

Note: For columns (1) and (2), the estimation sample consists of respondents who faced just one out of two modules (between-design group). For columns (3) and (4), the estimation sample consists of respondents who faced both the environmental and financial numeracy module (within-design group). Robust standard errors are in parentheses. Controls consist of the dummy variables: female, parent, working, married, very good or excellent health (above median), fair or poor health (below median), and the approximate quartile age groups 36-55 years, 56-66 years, and 66+ years (omitted age group is 35- years). \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## 15.4 Conclusion

In this study, we introduced new modules to proxy people's grasp of the long-term future by measuring their understanding of long-term exponential processes. These modules relate to previously used instruments in the literatures on financial literacy and exponential growth bias. Our preliminary results indicate a wide variety of respondent performance across questions, facilitating our proposed analysis.

We observe that performance tends to decline as the question's time horizon extends, suggesting either increased mathematical complexity or a general difficulty in conceptualizing the long-term future. Future work will delve deeper into these potential drivers. Additionally, the observed differences in performance patterns between the financial and environmental modules indicate that pre-existing attitudes towards intergenerational challenges may also influence outcomes.

Our work explores a potential barrier to confronting long-term societal challenges. Intergenerational time preferences are an important indicator for policy design, and we aim to investigate the role that people's detachment from the long-term future might play. Our first results suggest that potential incentives and educational programs should not only address people's misperceptions of long-term processes but also their additional contextual

biases. Moreover, we believe that our question modules are valuable for future research, given their ability to capture how context influences understanding of long-term processes.

## **15.5 References**

- Bradford, W. D., Dolan, P., & Galizzi, M. M. (2019). Looking ahead: Subjective time perception and individual discounting. *Journal of Risk and Uncertainty*, *58*(1), 43–69.
- Cropper, M. L., Aydede, S. K., & Portney, P. R. (1994). Preferences for life saving programs: How the public discounts time and age. *Journal of Risk and Uncertainty*, *8*(3), 221–229.
- Frederick, S. (2003). Measuring intergenerational time preference: Are future lives valued less? *The Journal Of Risk And Uncertainty*, *26*(1), 39-53.
- Johannesson, M., & Johannsson, P. O. (1997). Saving Lives in the Present Versus Saving Lives in the Future - Is There a Framing Effect? *Journal of Risk and Uncertainty*, *15*(2), 167–176.
- Kaiser, T., & Lusardi, A. (2024). Financial literacy and financial education: An overview (National Bureau of Economic Research Working Paper No. 32355). National Bureau of Economic Research. <https://doi.org/10.3386/w32355>
- Kip Viscusi, W., & Zeckhauser, R. J. (2006). The perception and valuation of the risks of climate change: A rational and behavioral blend. *Climatic Change*, *77*, 151–177.
- Lammers, J., Crusius, J., & Gast, A. (2020). Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(28), 16264–16266.
- Levy, M., & Tasoff, J. (2016). Exponential-growth bias and lifecycle consumption. *Journal of the European Economic Association*, *14*(3), 545–583.



- Lusardi, A., & Mitchell, O.S. (2008). Planning and financial literacy: How do women fare? *American Economic Review*, *98*(2), 413–417.
- Stango, V., & Zinman, J. (2009). Exponential growth bias and household finance. *The Journal of Finance*, *64*(6), 2807–2849.
- Tonn, B., Hemrick, A., & Conrad, F. (2006). Cognitive representations of the future: Survey results. *Futures*, *38*(7), 810–829.
- Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449–472.
- Voorintholt, L. (2020). Discounting Beyond Death: An exploration of intergenerational distribution preferences. *Reinvention: an International Journal of Undergraduate Research*, *13*(1).
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416–422.
- University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## **16. Flexibility stigma in the UK post-pandemic : intersectionality of stigma across gender and parental lines**

*Heejung Chung, King's College London*

*Senhu Wang, National University of Singapore*

### **16.1 Introduction**

The pandemic has led to a rapid increase in the number of workers working from home, and hybrid working – namely where workers work from home a few days a week and the rest in their employer's premises. For example, up to half of the workforce across Europe worked from home during the peak of the pandemic, and even in 2021 up to 1/3 of employees across Europe were still doing some of their work from home (Eurofound, 2022). Prior to the pandemic, there was a belief that homeworkers are less productive, motivated, and committed compared to workers who come into the office (Chung, 2022; Chung & Seo, 2023; Williams et al., 2013). Such beliefs largely stem from the fact that homeworking makes workers deviate from the ideal worker standard that exists in many of our societies, expecting workers to prioritise work above all else in life and work long hours in the office (Williams, 1999). The sudden rise in homeworking, along with the fact that homeworking was not a request by workers but was generally imposed by governments by law to contain the virus, has resulted in significant changes in managers' attitudes in the UK (e.g., CIPD, 2021). Many studies during the pandemic and just after lockdown periods show that increasingly managers see the productivity and other potential gains homeworking can bring to the company (ONS, 2021). However, stigmatised views against homeworkers are steadily returning, with many managers asking workers to come back into the office citing productivity as a key issue (Partridge, 2023; The Economist, 2023).

When stigmatised ideas around homeworkers persist, workers are unlikely to take up homeworking arrangements due to fears of negative career penalties (Thébaud & Pedulla, 2022). This may be especially true for certain groups of the population, on one hand

because they are generally responsible for the breadwinning roles within households like men (Rudman & Mescher, 2013) or as they already suffer from negative bias against their work capacities – e.g. ethnic minority workers (Chung et al., 2024). Stigmatised views against flexible workers are also likely to result in workers working harder and longer when working from home, feeling a need to be always-on and connected (Chung, 2022), with blurred boundaries between work and family life resulting in work encroaching on private time (Glass & Noonan, 2016). This results in negative well-being outcomes for workers, and negative outcomes for companies with regards to decline in productivity, burn out, and other potential issues and costs (Isham et al., 2021; Kelly et al., 2014). In other words, we need to eliminate the stigmatised views against homeworkers to reap the multitude of benefits the rise of home and hybrid working can bring.

The survey content designed for the UKHLS aims to address these issues in five ways. Firstly, we aim to capture flexibility stigma in a more meaningful manner, using a large-scale representative sample of the population. Despite daily news headlines talking about managers' negative perceptions towards homeworkers (e.g., BBC, 2021), studies on flexibility stigma in the UK are lacking. What data exist mostly rely on surveys capturing respondents' own evaluations of the situation (e.g. British Social Attitude Survey 2018). As stigma tends to be unconscious, we need better ways of capturing these perceptions that individuals themselves may not be aware of. Following previous successful studies of the US (Brescoll et al., 2013; Munsch, 2016; Thébaud & Pedulla, 2022), we used experimental vignette survey designs to capture underlying stigma against flexible workers. Secondly, previous experimental studies generally use online convenience samples (Munsch, 2016). In contrast, this paper provides evidence in the UK using a large-scale representative sample which the UK Household Panel Survey provides.

In addition, through this survey we wanted to explore further for whom homeworking is more likely to lead to stigmatised views against them. Some scholars (e.g., Kelland et al., 2022; Rudman & Mescher, 2013) argue that men and fathers are more likely to experience flexibility stigma as flexible working make them deviate away more from their expected roles as providers of the family. Others (e.g., Brescoll et al., 2013; Chung, 2020) argue that

mothers are more likely to face stigma, as their ability and commitment to work is already questioned due to social norms that assume that mothers will (always) prioritise children and households above all else. The paper explores how the intersection of gender and parental status shape respondents' views of homeworkers' productivity, commitment and their potential for promotion (see also, Brescoll et al., 2013). Fourthly, we compare hybrid workers who mostly come into the office (working from home only 1-2 days a week) versus those who mostly work from home (3-4 days a week). Given the proximity bias which shapes stigmatised views against homeworkers (Cristea & Leonardi, 2019), we expect the number of days in which workers work from home to matter. What is more, we expect this may matter more for certain groups in the labour market more than others. Finally, we contribute to the debates around flexibility stigma by examining how changes in the organisational and national policy changes can help reduce the scarring effect homeworking can have for workers (see also, Kelly et al., 2014; Moen et al., 2016). More specifically, we examine how stigmatised views against homeworking may be reduced when more workers work from home regularly (see also, van der Lippe & Lippényi, 2020), and it is not framed as a policy for parents or mothers. This part is yet to be analysed and will not be presented in this chapter but we will present the theoretical assumptions behind the role of context in shaping stigmatised views against hybrid workers.

Normalisation of flexible working has been shown in many studies, including field experiment studies, to help shape attitudes towards flexible working, reducing the likelihood that flexible workers are stigmatised (Angelici & Profeta, 2020; Kelly et al., 2014; Munsch et al., 2014; van der Lippe & Lippényi, 2020). When more workers work from home, homeworkers' deviation away from the norm is reduced, reducing the likelihood that they will be singled out, stigmatised and experience career penalties. When there are policies to provide workers homeworking rights, homeworking may no longer be perceived as a gift that needs to be reciprocated, but a right (Been et al., 2017). The normalisation of homeworking, especially to all workers, may also signal that homeworking policies are performance enhancing, and/or are used based on employer-led motives, such as reduction of rental costs, rather than based on workers' initiatives alone (Anderson & Kelliher, 2020). This study explored normalisation through two means – whether formal policies on hybrid

working exist at the organisational level, and the extent to which hybrid working is a common practice within the company. We do not include policies at the national level as the UK has a right to request flexible working policy which covers all workers as of 2014.

When homeworking is attributed to parents and mothers, it may be because it is largely women and parents who take it up, or because policies are stated specifically for these groups, it is most likely to be perceived as being used (solely) for family-friendly purposes. This is because parents, especially mothers, are often perceived as having the most pressing need for balancing work and family responsibilities (Pedulla & Thébaud, 2015). In such cases, homeworking arrangements are assumed to not lead to improved performance (Leslie et al., 2012) despite the fact that balancing work and family responsibilities can actually enhance performance outcomes (Kelly et al., 2014). This assumption that homeworking is primarily a work-family arrangement may cause workers to deviate from the traditional ideal worker image, and as a result face more penalties (Leslie et al., 2012; Williams et al., 2013). In our experimental set up, we explore context variation of when there are only policies for/and hybrid working is used by mothers, or just parents, or all workers, compared to when no such policies/take up exists.

Using an experimental vignette survey setting, we aim to test to causal mechanisms of the impact of contexts. By eliminating other potential confounders, e.g. culture, that may be driving the variation, experimental settings help us better understand the impact of the context variations on stigma. The data we used was collected in the UK towards the end of 2023, a period when debates around the need for workers to return to office was widespread. Using the Innovation Panel of the UK Household Panel Survey (University of Essex, Institute for Social and Economic Research, 2024), we asked workers their thoughts on a fictitious worker's productivity, commitment, team playing nature and potential for promotion.

## **16.2 Methods**

### **Research design**

This study used a factorial survey experiment (FSE) including an online survey and a vignette experiment, which was carried out in Singapore. Combining the advantages of survey and experiment, FSEs have been increasingly used to better understand the multi-dimensionality of individuals' decision-making process (Auspurg et al., 2017; Shi & Wang, 2022). In this study, there are at least three advantages of using an FSE to study employer evaluations of teleworking. First, as the randomization procedure of the experimental design has eliminated confounding effects of all respondent features, the effects of employee teleworking status on employer ratings can be interpreted as causal effects. The complete control over respondent characteristics ensures a higher level of internal validity compared to traditional survey methods. Second, respondents in an FSE can be asked to evaluate a large number of vignettes, which vary in multiple dimensions. This has the advantage of allowing us to test for various combinations, including some which occur only rarely in reality (Auspurg & Hinz, 2014). Third, the multi-dimensionality of the FSE could help prevent respondents from giving socially desirable answers because the simultaneously changing vignette dimensions makes it difficult for them to know the study's purposes (McDonald, 2019).

### ***Vignette set up***

All respondents were given an introductory statement of a hypothetical promotions panel in which they are participating. They were then asked to provide their evaluation of three candidates who varied in their gender, parental status, hybrid working status, and organisational context. Note that caution has been taken when selecting male and female names so that they are of a generic worker – of non-ethnic minority background (generally white), and are generally classless. We have done this to ensure that there are no potential influence of class and racial bias captured through this vignette.

We have asked individuals to rate the candidate with regards to their perceived level of commitment, productivity, the extent to which they are a team player, and finally their suitability for the promotion.

**Table 16-1 Vignette experimental variables**

<b>Experimental variables</b>	<b>Levels</b>	<b>Descriptions</b>
Gender (2)	1. Male 2. Female	1. Different male names 2. Different female names
Parenthood (2)	1. No 2. Yes	1. have no child 2. have two children
Home/hybrid-working practice (3)	1. Works in the office 2. Homeworking 1-2 days a week 3. Homeworking 3-4 days a week	1. Works in the office 2. Works from home 1-2 days a week on a regular basis 3. Works from home 3-4 days a week on a regular basis
Organizational context 1 (4)	1. Low policy use 2. High policy use for mothers 3. High policy use for parents 4. High policy use for all workers	1. In the company, less than 20% of all workers work from home on a regular basis 2. In the company, more than half of all mothers work from home on a regular basis 3. In the company, more than half of all parents work from home on a regular basis 4. In the company, more than half of all workers work from home on a regular basis
Organisational context 2 (4)	1. No formal policy 2. Formal policy for mothers 3. Formal policy for parents 4. Formal policy for all workers	1. There are no company level policies on hybrid-working 2. There is a company level policy that allows hybrid-working only for mothers 3. There is a company level policy that allows hybrid-working only for parents 4. There is a company level policy that allows hybrid-working for all workers

***Set up text***

*Imagine you are a human resource manager in a company. The company has a job vacancy for a full-time open-ended contract and you are responsible for selecting a suitable candidate, who needs to be a decisive person with the ability to analyse information. They will lead a team of*

*around 10 people to maintain good relationships with customers and develop new markets. A number of internal candidates have been selected, all of whom have the proper experience and educational background. Now you will be shown three candidates. Please assess the suitability for promotion and work commitment for each candidate.*

***Candidate 1***

**Charlotte Davies (Gender)**, aged 35, **has no children (Parental status)**, holds a master’s degree in Economics and Management. She has worked from home **one or two days a week on a regular basis (homeworking status)** in the last year. **More than half of all workers (organisational context: normalisation)** in the company work from home on a regular basis and **there is a formal company level policy that allows for hybrid-working for parents (organisational context: gendered policies)**.

Q1. How likely is it that would you recommend Charlotte for the job?

0 (not at all likely) ----- 10 (most likely)

Q2. How do you see her in terms of her work commitment?

0 (very poor) ----- 10 (very good)

Q3. How do you see Charlotte in terms of her productivity?

0 (very poor) ----- 10 (very good)

Q4. How do you see her as a team player?

0 (very poor) ----- 10 (very good)

To test the respondents’ attention the following question was added.

Q5. Please recall, which piece of information has appeared in the candidate’s profiles?

Teleworking status



- Trade union membership status
- Sport club membership status

## Data and sample

Table 16-2 reports respondents' characteristics included in our data set.

**Table 16-2 Sample characteristics**

	<b>M, %</b>	<b>Min</b>	<b>Max</b>
Age, M (SD)	43.25 (13.13)	18	65
Age group, %			
18-25	11.74		
26-35	20.57		
36-45	21.74		
46-55	22.67		
56-65	23.29		
Gender, %			
Men	44.78		
Women	55.22		
Marital status, %			
Never married	37.76		
Married	50.50		
Divorced/separated/widowed	11.74		
Employment status, %			
Employed	84.54		
Not employed	15.46		
Longstanding illness, %			
Yes	28.68		
No	71.32		
Weekly work hours, M (SD)	28.28 (19.65)	0	100
Number of respondents	1,470		

Note. M = Means, % = Proportions, SD = Standard deviations.

## Analytic strategy

As the data are hierarchically structured (vignettes are embedded within the respondents) and dependent variables are treated as continuous, random intercept multilevel linear regression models were applied to yield correct standard errors. Due to the successful randomization in the experimental design, the assumption of exogeneity is satisfied and

therefore a multilevel model rather than fixed-effects model is used to maximize statistical efficiency. The multilevel model can be expressed in the following equation.  $Rating_{ij}$  is the dependent variable measuring the extent to which a manager rates the specific candidate's promotion opportunities (vignette  $i$  for respondent  $j$ ).  $Homeworking\ status_{ij}$  is the explanatory variable measuring different teleworking statuses of the vignette  $i$  for respondent  $j$ . In addition, there are a number of moderator variables including gender-parenthood status, organizational teleworking use and government teleworking policy. Finally,  $\alpha_{ij}$  is the random intercept;  $c_j$  is the respondent-level error term; and  $\mu_{ij}$  is the vignette-level error term.

$$Rating_{ij} = \alpha_{ij} + \beta_1 Homeworking\ status_{ij} + \beta_2 Gender\_parenthood\ status_{ij} + c_j + \mu_{ij}$$

Specifically, we first estimated the main model by investigating the effects of teleworking status on employer ratings. Next, we fitted interaction terms between teleworking status and gender-parenthood status to explore whether and how the effects of teleworking vary across demographic groups. Finally, we conducted mediation analyses to explore whether perceived work commitment, productivity or team spirit can explain the effects of homeworking on promotion opportunities.

### 16.3 Results

**Table 16-3 Random effects multilevel models examining the effects of homeworking on degree of promotion recommendation**

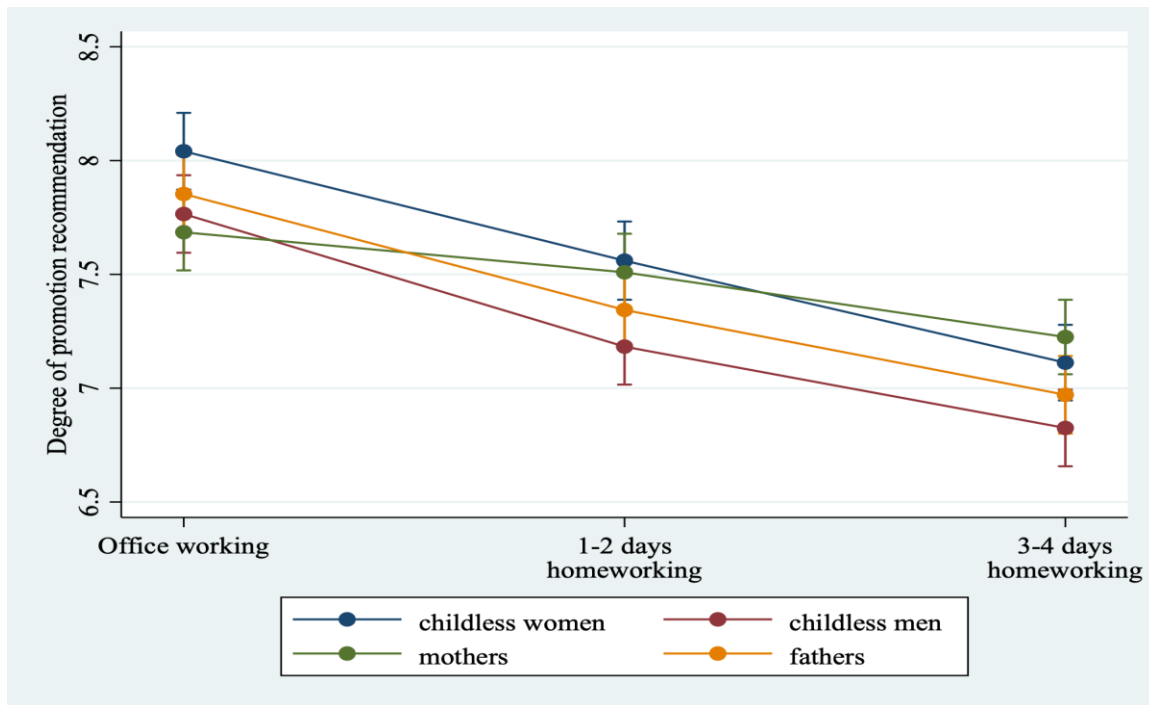
	Model 1	Model 2
Homeworking (ref. = No)		
1-2 days	-0.44*** (0.06)	-0.18 (0.11)
3-4 days	-0.80*** (0.05)	-0.46*** (0.11)
Gender-parenthood (ref. = Mothers)		
Childless women		0.36** (0.11)
Childless men		0.08 (0.11)
Fathers		0.17 (0.11)
Homeworking × Gender-parenthood		
1-2 days × Childless women		-0.30* (0.16)
1-2 days × Childless men		-0.41** (0.15)
1-2 days × Fathers		-0.33* (0.16)
3-4 days × Childless women		-0.47** (0.15)
3-4 days × Childless men		-0.48** (0.15)
3-4 days × Fathers		-0.42** (0.15)
Constant	7.84*** (0.05)	7.69*** (0.09)
R-squared	0.03	0.04
Number of vignette-respondent observations	4,320	4,320
Number of respondents	1,470	1,470

Note. Standard errors in parentheses. \*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Table 16-3 reports the results of random effects multilevel models. In model 1, we find that compared with those working in the office, those working from home 1-2 days ( $B = -0.44$ ,  $p < 0.001$ ) or 3-4 days ( $B = -0.80$ ,  $p < 0.001$ ) are significantly less likely to be recommended to promotion. Further analyses show that compared with 1-2 days homeworking, 3-4 days

homeworking has a significantly larger negative impact on the degree of promotion recommendation ( $B = -0.35, p < 0.001$ ). Overall, these results lend strong support to our hypothesis, suggesting a strong association between homeworking and flexibility stigma especially for a longer time of homeworking. In model 2 in Table 16-3, we further examined the interaction effect between gender-parenthood group and homeworking to explore whether the negative effect of homeworking varies across different gender and parenthood groups. All interaction effects are significant and negative. These results indicate that compared with mothers, the negative effects of homeworking regardless of its length are significantly more pronounced for childless women, childless men, and fathers. Figure 16-1 visualizes the interaction results by predicting the degree of promotion recommendation by homeworking status and gender-parenthood groups. It shows that among those working in the office, mothers have the lowest level of promotion recommendation compared with other groups. While homeworking especially of 3-4 days is associated lower degree of promotion recommendation for all groups, the slopes are much flatter for mothers than for the other three groups. These patterns indicate that mothers are less likely to be stigmatized compared with other groups due to their homeworking practice. Rather, they are likely to be stigmatised due to their motherhood status.

**Figure 16-1 Degree of promotion recommendation by homeworking status and gender-parenthood groups**



Next, in Table 16-4 we explore three possible mediating mechanisms (i.e., perceived commitment, productivity and team spirit) underlying the negative effects of homeworking for each gender-parenthood group. For childless women (in Panel A), the negative effects of 1-2 days or 3-4 days homeworking can be fully explained by their lower perceived commitment, productivity, and team spirit. It is important to point out that the mediation percentage for childless women working from home 1-2 days slightly exceeds 100%, indicating a minor suppression effect. This means that after accounting for mediators, the impact of homeworking shifts to a slightly positive direction, though it remains statistically insignificant. This can be partly attributed to respondents placing greater emphasis on the commitment, productivity, and team spirit of those working from home, compared to their promotion prospects, thereby introducing more variability into these factors. Consequently, we interpret the mediation percentage just over 100% as a reflection of random variation, suggesting full mediation. In Panel A, the mediation effects for all three mediators are significant, but commitment has larger mediation effects and can explain around 73% and 50% of the effects of 1-2 days and 3-4 days homeworking. In contrast, team spirit can explain around 30% and productivity can explain about 14%-20%.

For childless men in Panel B, we find that around 90% of the total effect can be mediated and all three factors are significant mediators, with the pattern being similar for 1-2 days or 3-4 days homeworking. Around 50% of the total effect of homeworking can be explained by lower work commitment. In contrast, team spirit and productivity can explain about 14-19% of the total effect.

For mothers in Panel C, there is full and slightly over-mediation of the total effect of homeworking for 3-4 days (similar to childless women) and all three mediators are significant. Work commitment again has the largest mediation percentage (59%), followed by productivity (47%) and then team spirit (27%). It is worth noting that the mediation percentage of productivity is much larger compared with other groups, suggesting that lower perceived productivity may be a particular concern for homeworking mothers.

For fathers in Panel D, around 73-80% of the total effect can be mediated and all three mediators are significant. The patterns for 1-2 days and 3-4 days homeworking are similar. Around 33-39% of total effect be explained by work commitment, followed by 25-27% of team spirit and then 16%-20% of productivity. We find that total mediation percentage for fathers is smaller compared with other gender and parenthood groups. And the importance of work commitment and productivity is also much weaker for fathers than for other groups especially mothers. This suggests that in addition to human capital other cultural factors such as violation of men's ideal worker norm may help explain the remaining effect.

**Table 16-4 Mediation effects of commitment, productivity, and team spirit**

	<b>1-2 days homeworking</b>	<b>3-4 days homeworking</b>
<i>Panel A. Childless women</i>		
Total mediation effect	Mediation effect = -0.65(0.16) *** % mediated = 125%	Mediation effect = -0.87(0.16) *** % mediated = 93%
Mediation effect of team spirit	Mediation effect = -0.17(0.04) *** % mediated = 32%	Mediation effect = -0.27(0.05) *** % mediated = 29%
Mediation effect of productivity	Mediation effect = -0.10(0.04) ** % mediated = 20%	Mediation effect = -0.14(0.05) ** % mediated = 14%
Mediation effect of commitment	Mediation effect = -0.38(0.08) *** % mediated = 73%	Mediation effect = -0.47(0.08) *** % mediated = 50%
<i>Panel B. Childless men</i>		
Total mediation effect	Mediation effect = -0.67(0.17) *** % mediated = 86%	Mediation effect = -0.89(0.18) *** % mediated = 89%
Mediation effect of team spirit	Mediation effect = -0.11(0.03) *** % mediated = 14%	Mediation effect = -0.19(0.04) *** % mediated = 19%
Mediation effect of productivity	Mediation effect = -0.14(0.04) *** % mediated = 18%	Mediation effect = -0.16(0.04) *** % mediated = 16%
Mediation effect of commitment	Mediation effect = -0.42(0.08) *** % mediated = 53%	Mediation effect = -0.54(0.09) *** % mediated = 54%
<i>Panel C. Mothers</i>		
Total mediation effect	NA (main effect non-significant)	Mediation effect = -0.65(0.15) *** % mediated = 133%
Mediation effect of team spirit	NA	Mediation effect = -0.13(0.04) *** % mediated = 27%
Mediation effect of productivity	NA	Mediation effect = -0.23(0.05) *** % mediated = 47%
Mediation effect of commitment	NA	Mediation effect = -0.29(0.06) *** % mediated = 59%
<i>Panel D. Fathers</i>		
Total mediation effect	Mediation effect = -0.33(0.16) * % mediated = 86%	Mediation effect = -0.56(0.16) *** % mediated = 73%
Mediation effect of team spirit	Mediation effect = -0.11(0.04) ** % mediated = 27%	Mediation effect = -0.19(0.05) *** % mediated = 25%
Mediation effect of productivity	Mediation effect = -0.08(0.03) * % mediated = 20%	Mediation effect = -0.12(0.04) ** % mediated = 16%
Mediation effect of commitment	Mediation effect = -0.15(0.06) * % mediated = 39%	Mediation effect = -0.25(0.06) *** % mediated = 33%

## **16.4 Conclusion**

This chapter examined the usefulness of a vignette survey approach in measuring extent to which there are stigmatised and biased views around homeworkers' productivity, commitment, their commitment to the team and how this can negatively influence their career outcome. The results show clear signs of negative career outcomes for homeworkers, especially those who work more days at home (3-4 days at home) compared to those who work in the office more often (1-2 days at home). Although all groups examined in our data – namely, mothers, fathers, men and women without children, experience flexibility stigma when homeworking, mothers' homeworking was met with the least bias. However, this was largely because mothers already experiencing bias due to their motherhood status, and homeworking not adding more to this bias against mothers. Much if not all of the decrease in promotion chances homeworkers face could be explained through the homeworker's perceived commitment, productivity of the homeworker and the homeworker's perceived engagement as a team player. On average, of the three, bias against workers' commitment levels seem to be most influential, followed by their willingness to be a team player. However, this varied depending on the type of worker. For example, mothers' expected decline in productivity was more important in understanding the bias they experience compared to other groups. This indicates that there may be more bias against mothers' capacity to focus on work when working from home compared to other groups of workers who may be expected to hold stronger boundaries between work and family. This is largely due to the assumed role mothers have in the household/childcare which is expected to hinder mothers' capacity to focus at work when at home (Chung, 2022; Chung & Van der Lippe, 2020; Parry, 2024).

The results of the study show the need for further policy interventions at both national and organisational levels to eliminate bias against flexible workers. This could include education and campaigns to help people, especially managers, better understand the productivity enhancing nature of homeworking and other work-family reconciliation practices (Beauregard & Henry, 2009; Bloom et al., 2015). Further policy interventions are needed to ensure that problematic gender normative views around whose role it is to care and do the



breadwinning is also needed. Finally, stronger rights for flexible workers including those that provide protective mechanisms against potential discrimination is needed based on the results of the study.

## **16.5 References**

- Anderson, D., & Kelliher, C. (2020). Enforced remote working and the work-life interface during lockdown. *Gender in Management: An International Journal*.
- Angelici, M., & Profeta, P. (2020). *Smart-working: Work flexibility without constraints*. Munich: Center for Economic Studies and ifo Institute (CESifo).
- Auspurg, K., & Hinz, T. (2014). *Factorial survey experiments* (Vol. 175). Sage Publications.
- Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, 82(1), 179-210.
- BBC. (2021). Goldman Sachs: Bank boss rejects work from home as the 'new normal'. *BBC News*. <https://www.bbc.co.uk/news/business-56192048> (accessed 26th of March 2021)
- Beauregard, T. A., & Henry, L. C. (2009). Making the link between work-life balance practices and organizational performance. *Human resource management review*, 19(1), 9-22.
- Been, W., den Dulk, L., & van der Lippe, T. (2017). A business case or social responsibility? How top managers' support for work-life arrangements relates to the national context. *Community, Work & Family*, 20(5), 573-599.
- Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics*, 130(1), 165-218.
- Brescoll, V. L., Glass, J., & Sedlovskaya, A. (2013). Ask and Ye Shall Receive? The Dynamics of Employer-Provided Flexible Work Options and the Need for Public Policy. *Journal of Social Issues*, 69(2), 367-388.

- Chung, H. (2020). Gender, flexibility stigma, and the perceived negative consequences of flexible working in the UK. *Social Indicators Research*, 151(2), 521-545.
- Chung, H. (2022). *The Flexibility Paradox: Why flexible working can lead to (self-)exploitation*. Policy Press.
- Chung, H., & Seo, H. (2023). Flexibility stigma - how national contexts can shift the extent to which flexible workers are penalised. *TransEuroWorks Working Paper Series*, 3/2023.
- Chung, H., & Van der Lippe, T. (2020). Flexible working work life balance and gender equality: Introduction. *Social Indicators Research*, 151(2), 365-381.  
<https://doi.org/DOI:10.1007/s11205-018-2025-x>
- Chung, H., Yuan, S., & Arkwright, A. (2024). *Making hybrid inclusive: Black workers experiences of hybrid working* London: Trades Union Congress.
- CIPD. (2021). *Flexible working: lessons from the pandemic* London: Chartered Institute for Personnel Development.  
<https://www.cipd.co.uk/knowledge/fundamentals/relations/flexible-working/flexible-working-lessons-pandemic#gref> (accessed 2nd April 2021)
- Cristea, I. C., & Leonardi, P. M. (2019). Get Noticed and Die Trying: Signals, Sacrifice, and the Production of Face Time in Distributed Work. *Organization Science*, 30(3), 552-572.
- Eurofound. (2022). *Working Conditions in the Time of COVID-19: Implications for the future* Luxembourg: Publications Office of the European Union.
- Glass, J. L., & Noonan, M. C. (2016). Telecommuting and Earnings Trajectories Among American Women and Men 1989-2008. *Social Forces*, 95(1), 217-250.
- Isham, A., Mair, S., & Jackson, T. (2021). Wellbeing and productivity: a review of the literature. *CUSP Working Paper Series*, No.22.
- Kelland, J., Lewis, D., & Fisher, V. (2022). Viewed with suspicion, considered idle and mocked-working caregiving fathers and fatherhood forfeits. *Gender, Work & Organization*, 29(5), 1578-1593.

- Kelly, E. L., Moen, P., Oakes, J. M., Fan, W., Okechukwu, C., Davis, K. D., Hammer, L. B., Kossek, E. E., King, R. B., Hanson, G. C., Mierzwa, F., & Casper, L. M. (2014). Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network. *American Sociological Review*, *79*(3), 485-516.  
<https://doi.org/10.1177/0003122414531435>
- Leslie, L. M., Manchester, C. F., Park, T.-Y., & Mehng, S. A. (2012). Flexible work practices: A source of career premiums or penalties? *Academy of Management Journal*, *55*(6), 1407-1428.
- McDonald, P. (2019). How factorial survey analysis improves our understanding of employer preferences. *Swiss Journal of Sociology*, *45*(2), 237-260.
- Moen, P., Kelly, E. L., Fan, W., Lee, S.-R., Almeida, D., Kossek, E. E., & Buxton, O. M. (2016). Does a flexibility/support organizational initiative improve high-tech employees' well-being? Evidence from the work, family, and health network. *American Sociological Review*, *81*(1), 134-164.
- Munsch, C. L. (2016). Flexible work, flexible penalties: the effect of gender, childcare, and type of request on the flexibility bias. *Social Forces*, *94*(4), 1567-1591.
- Munsch, C. L., Ridgeway, C. L., & Williams, J. (2014). Pluralistic ignorance and the flexibility bias: Understanding and mitigating flextime and flexplace bias at work. *Work and Occupations*, *41*(1), 40-62.
- ONS. (2021). *Business and individual attitudes towards the future of homeworking, UK: April to May 2021* London: Office for National Statistics.
- Parry, A. (2024). The flexibility paradox and spatial-temporal dimensions of COVID-19 remote work adaptation among dual-earner mothers and fathers. *Gender, Work & Organization*.
- Partridge, J. (2023, 5 Oct 2023). Two-thirds of CEOs think staff will return to office five days a week, survey finds. *The Guardian*.  
<https://www.theguardian.com/money/2023/oct/05/two-thirds-ceos-think-staff-return-to-office-five-days-a-week-survey-finds>

- Pedulla, D. S., & Thébaud, S. (2015). Can we finish the revolution? Gender, work-family ideals, and institutional constraint. *American Sociological Review*, *80*(1), 116-139.
- Rudman, L. A., & Mescher, K. (2013). Penalizing men who request a family leave: Is flexibility stigma a femininity stigma? *Journal of Social Issues*, *69*(2), 322-340.
- Shi, L. P., & Wang, S. (2022). Demand-side consequences of unemployment and horizontal skill mismatches across national contexts: An employer-based factorial survey experiment. *Social Science Research*, *104*, 102668.
- The Economist. (2023, June 28th, 2023). The working-from-home illusion fades. *The Economist*.
- Thébaud, S., & Pedulla, D. S. (2022). When do work-family policies work for men and women? Unpacking the effects of formal policies versus informal practices. *Work and Occupations*, *49*(2), 229-263.
- University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.
- van der Lippe, T., & Lippényi, Z. (2020). Beyond Formal Access: Organizational Context, Working From Home, and Work–Family Conflict of Men and Women in European Workplaces. *Social Indicators Research*, *151*(2), 383-402.  
<https://doi.org/10.1007/s11205-018-1993-1>
- Williams, J. (1999). *Unbending gender: Why family and work conflict and what to do about it*. Oxford University Press.
- Williams, J., Blair-Loy, M., & Berdahl, J. L. (2013). Cultural schemas, social class, and the flexibility stigma. *Journal of Social Issues*, *69*(2), 209-234.

## **17. The relationship between employment status, passage of time judgements and wellbeing**

*Sharon Raj, London School of Economics and Political Science*

*Liam Delaney, London School of Economics and Political Science*

### **17.1 Introduction**

The psychological harm caused by unemployment is one of the most robust findings in wellbeing research. Unemployment is one of the few life events that individuals do not fully adapt to (Clark et al., 2008; Clark & Georgellis, 2013; Clark, 2018) and, even after new work is found, individuals can be left permanently scarred with wellbeing settling at lower levels than that of those who have never been unemployed (Clark et al, 2001; Lucas et al, 2004). Much research has been conducted to explore the mechanisms underpinning these findings, but the underlying causes are still not fully understood. One specific aspect of the experience of unemployment where very little quantitative research has been conducted is how those who are unemployed experience time. While information on how people organise and use their time is available, no major panel survey to date has gathered information on how people perceive, or experience, their time.

Thus, while research on time perception is a growing field, the work conducted to date has tended to employ relatively small samples and has focused mainly on very short-term prospective or retrospective perceptions of time duration (Tanaka & Yotsumoto, 2017; Ogden, 2020), or on how passage of time judgements are affected by attention or emotion in the moment (see, for example, Droit-Volet & Wearden, 2015; Droit-Volet & Wearden, 2016; Droit-Volet et al., 2017; Rankin et al., 2019). Research exploring whether time perceptions are affected by less transitory individual circumstances is more limited, and we are not aware of any quantitative work specifically focused on the impact of unemployment. With this in mind, we included five questions concerning passage of time judgements in IP16 (University of Essex, Institute for Social and Economic Research, 2024) so that we could explore if, and how, one aspect of time perception – passage of time judgements – varied by

employment status and, if it did, whether such differences mediated levels of reported life satisfaction and general mental health.

## **17.2 Methods**

We included the five questions shown in Figure 17-1 in IP16. These questions had previously been used together by Wittmann & Lehnhoff (2005) when investigating age effects in time perception.<sup>9</sup> Participants responded to each question using a 5-item scale ranging from “*very slowly*” to “*very fast*”.

**Figure 17-1 Passage of Time Questions included in IP16**

- **How fast does time usually pass for you?**
- How fast do you expect the next hour to pass?
- How fast did the previous week pass for you?
- How fast did the previous month pass for you?
- How fast did the previous year pass for you?

We used the responses gathered to explore two core research questions:

- RQ1: Are passage of time judgements different between the full-time employed and the unemployed and, if so, how?
- RQ2: If a difference in passage of time judgements exists, does this help to explain the significant differences in life satisfaction and general mental health that are commonly seen between the full-time employed and the unemployed?

To address RQ1 we used OLS regression analysis to compare the responses to each of the five passage of time questions given by i) those who reported that they were in paid employment and working for 35 or more hours per week; and ii) those who reported that

---

<sup>9</sup> The first question was referenced by Wittmann & Lehnhoff as being used by Baum et al. (1984) and Staudinger et al. (1999) and the second to Richter & Benzenhofer (1985).

they were unemployed.<sup>10</sup> Since there were significant differences in the demographic characteristics between these groups, we controlled for age, gender, whether the individual had a partner sharing their home and the number of children under the age of 15 within the household in each of the regressions. In addition, we used information provided by the survey respondents regarding the reason for their unemployment to create a dummy variable to represent whether the cause of unemployment was endogenous or exogenous. This was also included as a control variable, helping to address a further source of potential confounding.

To address RQ2 we employed the same two comparison employment groups (full-time employed and unemployed) and identical control variables throughout. We created a single score of general mental health for each participant from the 12 separate General Health Questionnaire questions that were included in IP16 and, as a first step, we regressed both this score and, separately, life satisfaction scores on employment status using OLS regression.<sup>11</sup> This was done to ensure that the expected pattern of lower mental health and worse life satisfaction among the unemployed relative to the full-time employed was evident in the sample. After confirming that this was the case, we created a blended time perception measure from the five separate passage of time measures. We conducted causal mediation analysis using the `MEDIATE` function within Stata 18, treating each of life satisfaction and our single GHQ score in turn as dependent variables, employment status as the independent variable, and the blended time perception measure as a potential mediator

---

<sup>10</sup> OLS regressions were used for ease of presentation and comparison. However, ordered logit regressions were also run for each of the research questions and the results are available on request from the authors. The ordered logit regression results were consistent with the OLS results.

<sup>11</sup> The GHQ score was derived as follows: each of the 12 individual GHQ questions was answered using a 1-4 scale. These individual scores were recalibrated using the bimodal GHQ scoring method (where original scores 1 and 2 are assigned a final score of 0; and scores 3 and 4 are assigned a final score of 1). The final scores were then summed to result in a total GHQ score that ranged from 0-12. An alternative GHQ score was also generated by simply summing the 12 individual 1-4 scores. All OLS regressions and the causal mediation analysis were repeated using this alternative measure and the results were unchanged.

to explore the extent to which the time perceptions of the unemployed may service as a pathway through which unemployment affects wellbeing.

### 17.3 Results

Descriptive Statistics for the Dependent Variables Split by Employment Status are shown in Table 17-1.

**Table 17-1 Descriptive Statistics for the Dependent Variables Split by Employment Status**

	Employed (n = 869)				Unemployed (n = 136)			
	Mean	S. D.	Range	N	Mean	S.D.	Range	N
POTJ: Usual speed of time passing	3.69	0.76	1-5	865	3.31	0.91	1-5	125
POTJ: Expectation for next Hour	3.50	0.81	1-5	864	3.23	0.92	1-5	123
POTJ: Past Week	3.79	0.88	1-5	865	3.50	1.04	1-5	125
POTJ: Past Month	3.87	0.83	1-5	864	3.48	1.05	1-5	126
POTJ: Past Year	3.94	0.84	1-5	865	3.64	1.04	1-5	127
GHQ Score (13-point scale)	1.87	3.05	0-12	869	3.53	4.61	0-12	136
Life Satisfaction (7-point scale)	5.13	1.35	1-7	857	4.24	1.74	1-7	128

Notes: The GHQ score was derived as follows: each of the 12 individual GHQ questions was answered using a 1-4 scale. These individual scores were recalibrated using the bimodal GHQ scoring method common in clinical research (where original scores 1 and 2 are assigned a final score of 0; and scores 3 and 4 are assigned a final score of 1). The final scores were then summed to result in a total GHQ score that ranged from 0-12. A higher GHQ score indicates poorer mental health. Life Satisfaction was scored using a 7-point scale. A higher Life Satisfaction score indicates greater satisfaction. All passage of time judgements (POTJ) were scored using a 5-item scale. A higher POTJ score indicates that time is perceived to be passing more quickly.



The results of the regression analysis conducted to address RQ1 are shown in Table 17-2. The unemployed responded that time passed more slowly than the full-time employed did for each of the five survey questions and, for all five questions, the differences were sizable and statistically significant at the 0.1% level.

**Table 17-2 The Relationship Between Employment Status and Passage of Time Judgements**

Dependent Variable:	Passage of Time Judgements				
	Usual Speed of time	Next Hour Expectations	POTJ : Past Week	POTJ : Past Month	POTJ : Past Year
Employment status (1=unemployed)	-0.41*** (0.09)	-0.35*** (0.09)	-0.40*** (0.11)	-0.47*** (0.11)	-0.41*** (0.11)
Constant	3.55*** (0.35)	3.56*** (0.37)	3.92*** (0.32)	4.18*** (0.33)	4.30*** (0.29)
Controls included	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.03	0.01	0.02	0.03	0.02
Sample size	978	975	978	978	980

Notes: Robust standard errors (in parentheses). Controls included: age, gender, whether participant has a partner sharing their home, the number of children aged 0-15 living at home and whether the reason for unemployment was endogenous or exogenous. Endogenous unemployment was defined as unemployment resulting from resignation, dismissal, or any unemployment where the participant refused to give a reason or selected the reason as “other” within the Understanding Society questionnaire. \*\*\*Significant at the 0.1% level; \*\*Significant at the 1% level; \*Significant at the 5% level.

Prior to addressing RQ2, we conducted OLS regression analysis to confirm that the lower levels of life satisfaction and poorer general mental health (as measured by the General Health Questionnaire scores) that are typically the case for the unemployed relative to full-time employed would be true for this sample. This was indeed the case, as shown in Table 17-3. For both measures, the differences were sizeable and statistically significant at the 0.1% level.

**Table 17-3 The Relationship Between Employment Status and Wellbeing**

Dependent Variable:	Measures of Wellbeing	
	GHQ12	Life Satisfaction
Employ. status (1=unemployed)	1.79*** (0.47)	-0.93*** (0.18)
Constant	-0.11 (0.79)	5.38*** (0.49)
Controls included	Yes	Yes
Adjusted R-squared	0.04	0.05
Sample size	993	973

Notes: Robust standard errors (in parentheses). Controls included: age, gender, whether participant has a partner sharing their home, the number of children aged 0-15 living at home and whether the reason for unemployment was endogenous or exogenous. Endogenous unemployment was defined as unemployment resulting from resignation, dismissal, or any unemployment where the participant refused to give a reason or selected the reason as “other” within the Understanding Society questionnaire. The GHQ-12 score was derived from a 12-item questionnaire. The response to each question was initially provided using a 4-point scale. These scores were recalibrated using the bimodal GHQ scoring method (0-0-1-1), to result in a total score that ranged from 0-12. A higher score indicates poorer mental health. Life Satisfaction was scored using a 7-item Likert scale. A higher score indicates greater satisfaction. \*\*\*Significant at the 0.1% level; \*\*Significant at the 1% level; \*Significant at the 5% level.

The results of our causal mediation analysis are shown in Table 17-4. For both general mental health and life satisfaction, time perception (as measured by passage of time judgements) served as a partial mediator. Post-regression analysis shows that the proportion of the total effect due to mediation was 22% ( $p = 0.035$ ) for the GHQ score and 17% ( $p = 0.003$ ) for life satisfaction.

**Table 17-4 Causal Mediation Analysis: The Relationship Between Wellbeing, Employment Status & Time Perception**

<b>Dependent Variable:</b> <b>GHQ</b>	<b>Coefficient</b>	<b>Robust SE</b>	<b>P-value</b>	<b>95% Confidence Interval</b>	
<b>Direct Effect</b> (Emp Status 1 vs 0)	0.997	(0.451)	0.027	0.114	1.880
<b>Indirect Effect</b> (Emp Status 1 vs 0)	0.279	(0.104)	0.007	0.075	0.482
<b>Total Effect</b> (Emp Status 1 vs 0)	1.276	(0.455)	0.005	0.384	2.168
<b>Dependent Variable:</b> <b>Life Satisfaction</b>	<b>Coefficient</b>	<b>Robust SE</b>	<b>P-value</b>	<b>95% Confidence Interval</b>	
<b>Direct Effect</b> (Emp Status 1 vs 0)	-0.723	(0.180)	0.000	-1.076	-0.370
<b>Indirect Effect</b> (Emp Status 1 vs 0)	-0.148	(0.046)	0.001	-0.238	-0.058
<b>Total Effect</b> (Emp Status 1 vs 0)	-0.871	(0.184)	0.000	-1.231	-0.511

Notes: Causal mediation analysis treated each of the single GHQ score and life satisfaction in turn as dependent variables, employment status as the independent variable and a blended measure of time perception measure as a potential mediator. Controls used throughout the analysis: age, gender, whether participant has a partner sharing their home, the number of children aged 0-15 living at home and whether the reason for unemployment was endogenous or exogenous. Endogenous unemployment was defined as unemployment resulting from resignation, dismissal, or any unemployment where the participant refused to give a reason or selected the reason as “other” within the Understanding Society questionnaire. The GHQ-12 score was derived from a 12-item questionnaire. The response to each question was initially provided using a 4-point scale. These scores were recalibrated using the bimodal GHQ scoring method (0-0-1-1), to result in a total score that ranged from 0-12. A higher score indicates poorer mental health. Life Satisfaction is scored using a 7-item Likert scale. A higher score indicates greater satisfaction. The single measure of time perception was generated by summing each of the scores from the 5 individual passage of time judgement questions.

## **17.4 Conclusion**

Our preliminary research indicates that the passage of time judgements of the unemployed may be materially different from those in full-time employment, with time feeling like it is passing far more slowly, both in the moment, and when periods of time in the past are considered retrospectively. Our work also suggests that this difference may partly help to explain the differences in levels of life satisfaction and general mental health that are typically found between these two groups in the wellbeing literature. However, further analysis using alternative regression approaches and additional control variables are necessary to test the robustness of these preliminary findings before firmer conclusions can be drawn.

Separately, we note that the inclusion of passage of time judgements in IP16 also allows for a range of other work examining the extent to which time perception is impacted by, and is predictive of, social and economic circumstances.

## **17.5 References**

- Clark, A. (2018). Four decades of the economics of happiness: Where next? Review of Income & Wealth, 64(2), 245–269. <https://doi.org/10.1111/roiw.12369>
- Clark, A., Diener, E., Georgellis, Y., & Lucas, R. (2008). Lags and leads in life satisfaction: A test of the baseline hypothesis. The Economic Journal, 118(529), F222–F243. <https://doi.org/10.1111/j.1468-0297.2008.02150.x>
- Clark, A., & Georgellis, Y. (2013). Back to baseline in Britain: Adaptation in the British Household Panel Survey. Economica, 80(319), 496–512. <https://doi.org/10.1111/ecca.12007>
- Clark, A., Georgellis, Y., & Sanfey, P. (2001). Scarring: The psychological impact of past unemployment. Economica, 68(270), 221–241. <https://doi.org/10.1111/1468-0335.00243>

- Droit-Volet, S., Trahanias, P., & Maniadakis, M. (2017). Passage of time judgments in everyday life are not related to duration judgments except for long durations of several minutes. *Acta Psychologica*, 173, 116–121.  
<https://doi.org/10.1016/j.actpsy.2016.12.010>
- Droit-Volet, S., & Wearden, J. H. (2015). Experience Sampling Methodology reveals similarities in the experience of passage of time in young and elderly adults. *Acta Psychologica*, 156, 77–82. <https://doi.org/10.1016/j.actpsy.2015.01.006>
- Droit-Volet, S., & Wearden, J. (2016). Passage of time judgments are not duration judgments: Evidence from a study using experience sampling methodology. *Frontiers in Psychology*, 7, 176. <https://doi.org/10.3389/fpsyg.2016.00176>
- Lucas, R., Clark, A., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science*, 15(1), 8–13. <https://doi.org/10.1111/j.0963-7214.2004.01501002.x>
- Ogden, R. (2020). The passage of time during the UK Covid-19 lockdown. *PLOS ONE*, 15(7), e0235871. <https://doi.org/10.1371/journal.pone.0235871>
- Rankin, K., Sweeny, K., & Xu, S. (2019). Associations between subjective time perception and well-being during stressful waiting periods. *Stress and Health*, 35(4), 549–559.  
<https://doi.org/10.1002/smi.2888>
- Tanaka, R., & Yotsumoto, Y. (2017). Passage of time judgments is relative to temporal expectation. *Frontiers in Psychology*, 8, 187.  
<https://doi.org/10.3389/fpsyg.2017.00187>
- University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.
- Wittmann, M., & Lehnhoff, S. (2005). Age effects in perception of time. *Psychological Reports*, 97(3), 921–935. <https://doi.org/10.2466/pr0.97.3.921-935>

## **18. Cognitive reflection and politically motivated reasoning**

*Chris Butler, University of Antwerp*

*Ceri Fowler, University of Oxford*

### **18.1 Introduction**

Motivated reasoning in general refers to the phenomenon of individuals processing information in such a way as to activate certain goals. In psychology, most research has been interested in when individuals process information with the aim of activating ‘directional goals’ rather than processing information with the aim of coming to an accurate judgement (Kunda, 1990). In recent years, a large body of research has found evidence that citizens are prone to engage in politically motivated reasoning (Cohen, 2003; Taber and Lodge, 2016; Ditto et al, 2019), a phenomenon where people deviate from rational judgement by processing information in a way that reinforces their existing support for a political party or movement (Taber and Lodge, 2006). One of the most consistent findings of evidence of politically motivated reasoning is that citizens judge the merits of policies differently depending on whether they are framed as being promoted by a particular group (Cohen et al., 2003; Ditto et al., 2019).

However, there remains a lack of scholarly consensus as to whether politically motivated judgements are examples of unconscious or deliberative thinking, which is important for understanding how individuals can avoid these deviations from rational judgement.

Previously, it was largely assumed that this kind of reasoning was a manifestation of fast, associative ‘Type I’ processing of information (Kahneman, 2011), whereby individuals rely on automatic, unconscious reasoning based on existing patterns or experiences.

A decade ago, Kahan (2013) challenged this conception of politically motivated reasoning by testing whether individuals who scored more highly in cognitive reflection tests (CRT) - designed to test the presence of ‘Type II’ thinking whereby individuals reason more carefully before responding to information – were in fact more likely to engage in politically motivated reasoning. He found that individuals who scored higher on the CRT were more

likely to dismiss the test as a poor indicator of open-mindedness if told that those who disagreed with them politically scored higher on the test. The implication of Kahan's finding is that individuals who demonstrated a capacity for such deliberative thinking were more likely to process information about the CRT in a way which reinforced the validity of their political identity, i.e. that this was a deliberative rather than instinctive response.

However, several follow-up studies failed to find such a link (Stromback et al., 2021; Tappin et al., 2021; Maguire et al., 2022; Baker et al., 2023; Stagnaro et al., 2023). These differing results can partly be explained by the different ways in which politically motivated reasoning and cognitive reflection have been measured. For example, Stromback et al. (2021), Baker et al. (2023) and Stagnaro et al. (2023) measure *numerical ability* rather than cognitive reflection.

Further, existing studies have not measured politically motivated reasoning in the way that Cohen et al. (2003) captured it – whether individuals are more likely to support a policy if told that it is favoured by their political in-group or out-group. Instead, studies in this field often infer motivated reasoning by whether or not individuals express support for a policy that is likely to be consistent with their in-group, such as Democrats being more likely to believe in climate change (Kahan, 2013), or individuals who favour equality over liberty being more in favour of gender quotas (Stromback et al., 2021).

In this study we return to the question of whether politically motivated reasoning is a manifestation of Type I or Type II thinking, but build on Kahan's (2013) design in four ways. Firstly, we operationalise politically motivated reasoning as where individuals' support for a policy is dependent on their preferred group's support for a policy. Secondly, rather than presuming the policies that different groups will support, we apply group frames to a neutral policy. Thirdly, we avoid the use of misinformation in our vignettes. Fourthly, we make use of panel data to avoid priming respondents for their political group identity.

## 18.2 *Methods*

The UK offers a good setting to test for the effects of politically motivated reasoning. Whereas public opinion in the USA has become polarised around partisan identity, there is considerable evidence that a similar phenomenon is taking place in the UK around Brexit referendum identities, with several studies finding evidence of politically motivated reasoning in line with Brexit group identities (Greene et al., 2021; Sorace and Hobolt, 2021; Sumner et al., 2023).

To test the link between cognitive reflection and motivated reasoning, we designed a study to test whether individuals who scored higher on a CRT test were more likely to display motivated reasoning when judging the merits of a policy proposal. Firstly, we wanted to identify a salient policy proposal for which there would be no prior expectations as to it being more supported by supporters or opponents of the UK's decision to leave the EU. To do so, in early 2022 we browsed the UK Parliament's petition website for issues which had attracted over 100,000 signatures, on which a Remain or Leave frame could be credibly applied, and identified "Make verified ID a requirement for opening a social media account".

We fielded a pre-registered experiment

([https://osf.io/dxfcu/?view\\_only=d7ba16ae75114a4486768e4db14698b4](https://osf.io/dxfcu/?view_only=d7ba16ae75114a4486768e4db14698b4)) in the summer 2023 wave of the Understanding Society Innovation Panel (University of Essex, Institute for Social and Economic Research, 2024), a long-running panel study of public opinion in the UK. First, participants were asked to complete a 3-item CRT. Due to over-use of the original CRT, we use the CRT-2 test developed by Thomson and Oppenheimer (2016) which significantly predicts performance on heuristics and biases tasks.

CRT-2 consists of the following questions:

**1. *If you're running a race and you pass the person in second place, what place are you in?***

(Correct answer: second. Intuitive answer: first)

**2. *A farmer had 15 sheep and all but 8 died. How many are left?***

(Correct answer: 8. Intuitive answer: 7)



**3. *Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?***

(Correct answer: Emily. Intuitive answer: June)

Respondents were then randomly assigned to one of three experimental conditions. In the control frame, they were just shown some information about the petition and the arguments in favour of it. In the remain and leave frames respondents were prompted that the petition had attracted particular support in areas that voted for Remain or Leave in the Brexit referendum. We were careful to word the vignettes in such a way that respondents were not misinformed; hence the vignettes only refer to the petition being popular in areas of higher Remain or Leave support rather than making a false explicit claim that the petition was more popular among Remain or Leave voters. Table 18-1 shows the exact wording of the treatment.

**Table 18-1 Experiment vignettes wording**

Control group wording	'Remain frame' wording	'Leave frame' wording
Over 170,000 people have signed a petition to make verified ID a requirement for opening a social media account. The creators of the petition argue that removing the option of anonymous accounts would reduce cyber-bullying and online abuse.	Over 170,000 people have signed a petition to make verified ID a requirement for opening a social media account. The creators of the petition argue that removing the option of anonymous accounts would reduce cyber-bullying and online abuse.  This petition was popular in areas that voted Remain in the Brexit referendum, which may be due to Remain voters believing that they suffer more abuse online.	Over 170,000 people have signed a petition to make verified ID a requirement for opening a social media account. The creators of the petition argue that removing the option of anonymous accounts would reduce cyber-bullying and online abuse.  This petition was popular in areas that voted Leave in the Brexit referendum, which may be due to Leave voters believing that they suffer more abuse online.

Then all participants were asked: ***To what extent would you support or oppose this proposal?***

We test the theory that politically motivated reasoning is more common among those who score higher in CRT tests using an ordered logistic regression, with respondents' evaluations of the policy proposal as the dependent variable, and a 3-way interaction of their Brexit identity, CRT score, and the experimental component as the independent variables. More specifically, we look for evidence in support of the following hypotheses:

H1a: Remain supporters who score higher on the CRT will be more supportive of the proposal if prompted that the petition was more popular in Remain-voting areas

H1b: Remain supporters who score higher on the CRT will be less supportive of the proposal if prompted that the petition was more popular in Leave-voting areas

H1c: Leave supporters who score higher on the CRT will be more supportive of the proposal if prompted that the petition was more popular in Leave-voting areas

H1d: Leave supporters who score higher on the CRT will be less supportive of the proposal if prompted that the petition was more popular in Remain-voting areas

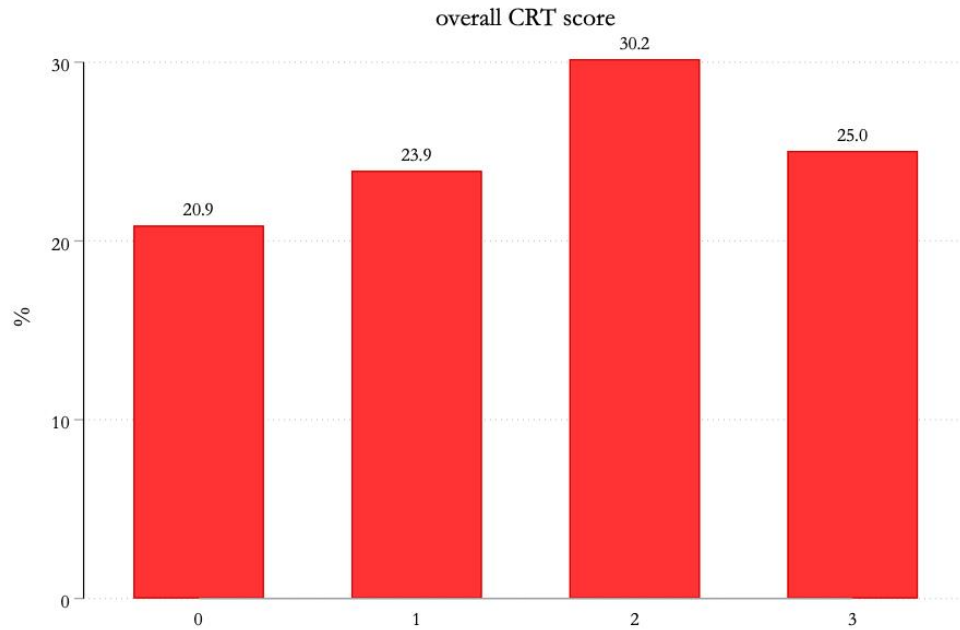
Brexit preference was captured by two later questions that asked respondents whether they voted in the European Union Referendum in June 2016 and how they voted. *Note: In subsequent analysis we will use past responses on referendum voting as the main operationalisation of this variable.*

### **18.3 Initial results**

#### **Descriptives**

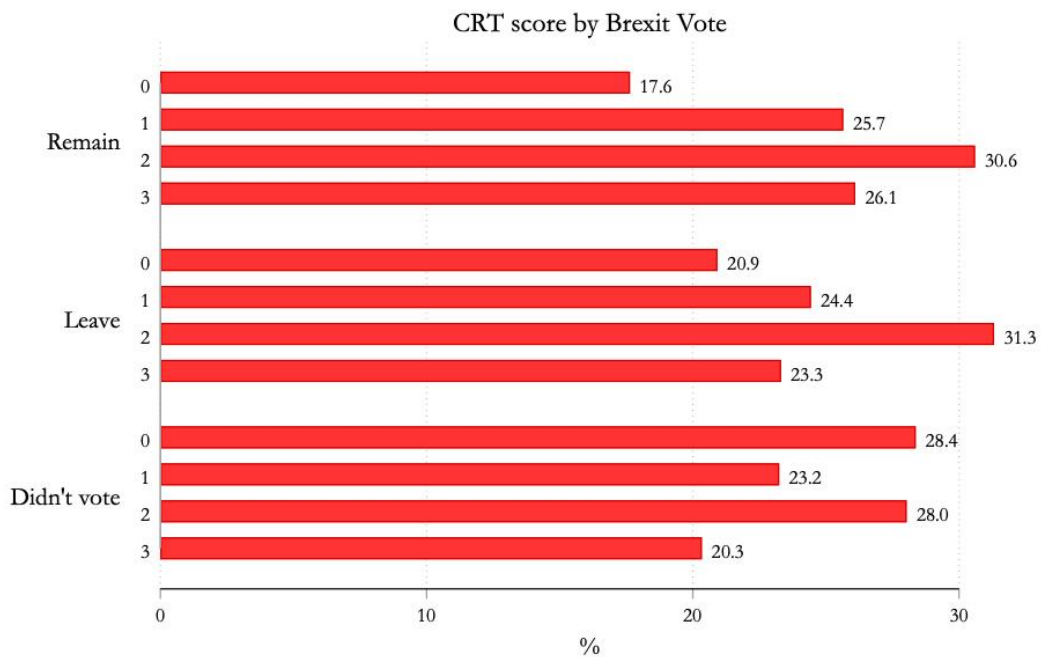
We start by examining how well respondents answered the 3 CRT questions. The aggregated CRT scores were created by simply summing an individual's three answers – if they give 3 correct answers they are awarded the maximum of 3, if they give 2 they are awarded a 2, etc. As can be seen in Figure 18-1 below, the scores were well spread, and whilst almost 80% of respondents managed to answer at least one question correctly, only 25% of respondents answered all three correctly.

**Figure 18-1 CRT scores overall. N=2840. Source: University of Essex, Institute for Social and Economic Research, 2024**



Next, we move on to examine whether respondents varied in their CRT score by their vote at the 2016 Brexit Referendum. Leave and Remain voters achieved similar scores, with the majority getting at least two questions correct.

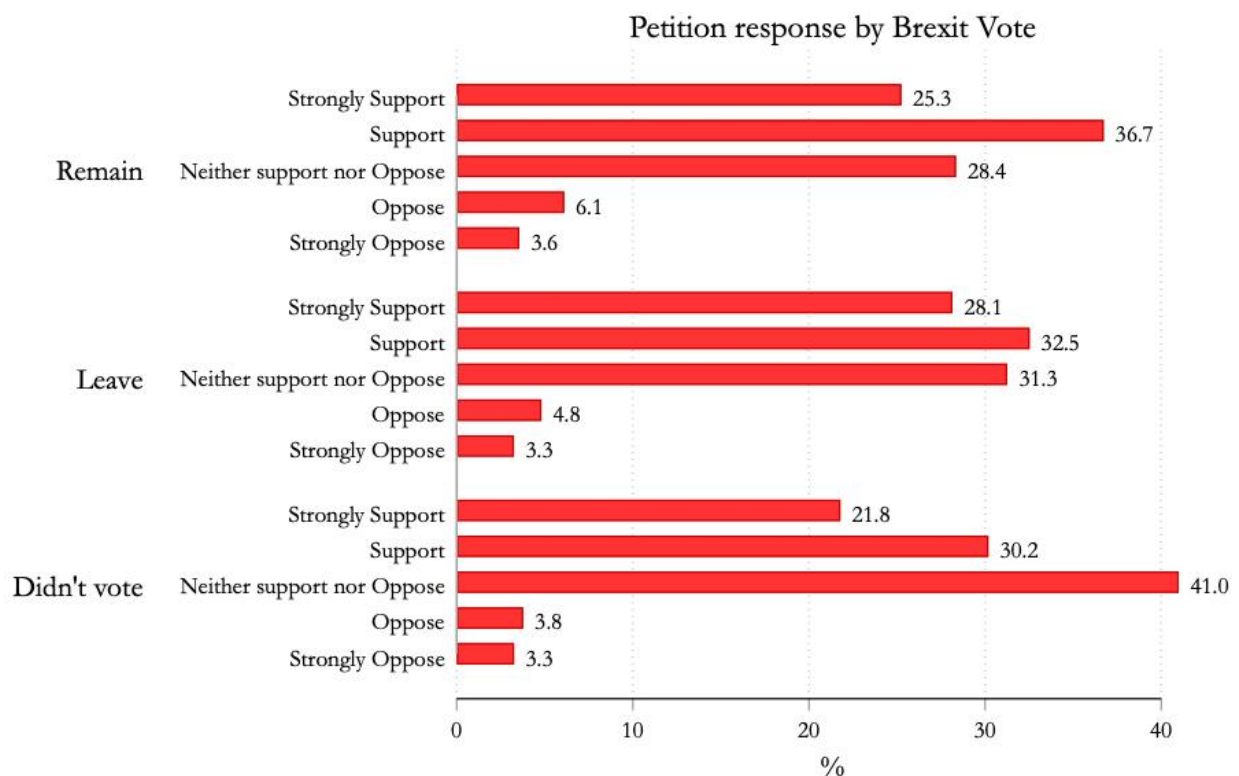
**Figure 18-2 CRT score by Brexit vote. N=2840. Source: University of Essex, Institute for Social and Economic Research, 2024**



However, among those who did not vote, scores were considerably lower, with over 28% of respondents failing to get at least one question correct. This may be as both CRT scores and turnout are correlated with education level, so indicates that controlling for education in the full model of support for the policy proposal will be important. Alternatively, it could indicate a general disengagement/apathy amongst non-voters which plays out both in the engagement with the survey when asked and in political behaviour.

Finally, we examine how supportive respondents were of the proposal to require ID to open a social media account, by Brexit vote choice, in Figure 18-3 below. Importantly for the later manipulation, results do not vary considerably between “Leavers” and “Remainers”, with the overall % who support the proposal being very similar. Very few individuals were unsupportive of the proposal. Those who didn’t vote were particularly likely to be neutral on the issue; this perhaps again indicates a general disengagement with political issues/the survey.

**Figure 18-3 Petition responses by Brexit Vote. N=2840. Source: University of Essex, Institute for Social and Economic Research, 2024**



## **Experimental results**

This section now presents the results of the main initial analysis. Table 18-2 presents the full results of the ordinal logistical regression, where support for the proposal is the dependent variable, and CRT score, Brexit vote, and version of the experiment the individual was presented with are the independent variables. We perform a 3-way interaction between the three independent variables to understand how partisans with different CRT scores responded to the different versions of the experiment.

***N.B. This data is not weighted and controls such as demographics and speed of response rate have not been added. Results should be treated as preliminary and subject to change when further analysed.***

**Table 18-2 Ordinal Logistic Regression with support for proposal as the outcome. N=2354.**

	Coefficient	Std. err.	Sig.
<b>CRT_score (0)</b>			
1	0.0	0.3	
2	-0.3	0.3	
3	0.1	0.3	
<b>brexitvote (Remain)</b>			
Leave	-0.4	0.4	
Didn't vote	-0.1	0.3	
<b>treatment (Control)</b>			
Remain Version	0.2	0.3	
Leave Version	0.8	0.3	*
<b>brexitvote#treatment</b>			
Leave#Remain Version	0.8	0.5	
Leave#Leave Version	0.1	0.5	
Didn't vote#Remain Version	0.1	0.5	
Didn't vote#Leave Version	-0.0	0.5	
<b>CRT_score#brexitvote#treatment</b>			
1#Leave#Remain Version	-0.4	0.7	
1#Leave#Leave Version	0.3	0.7	
1#Didn't vote#Remain Version	0.3	0.7	
1#Didn't vote#Leave Version	0.8	0.7	
2#Leave#Remain Version	-1.0	0.6	
2#Leave#Leave Version	-0.5	0.6	
2#Didn't vote#Remain Version	-0.1	0.7	
2#Didn't vote#Leave Version	-0.4	0.6	
3#Leave#Remain Version	-1.2	0.7	
3#Leave#Leave Version	-0.9	0.7	
3#Didn't vote#Remain Version	-1.4	0.7	
3#Didn't vote#Leave Version	-0.8	0.7	

Table 18-2 shows largely insignificant results, with the exception of receiving the Leave version of the treatment. Compared to the control group, those who received the Leave version of the treatment were significantly more likely to support the proposal than the

control group; those who received the Remain version were also more likely to more likely to support the proposal but the result was non-significant. The 2-way interaction between Brexit vote and treatment shows that the greater support for the policy with the Leave treatment was not driven by Leave voters. The change in an individual's likelihood of supporting the proposal did not vary significantly by their CRT score or by Brexit vote.

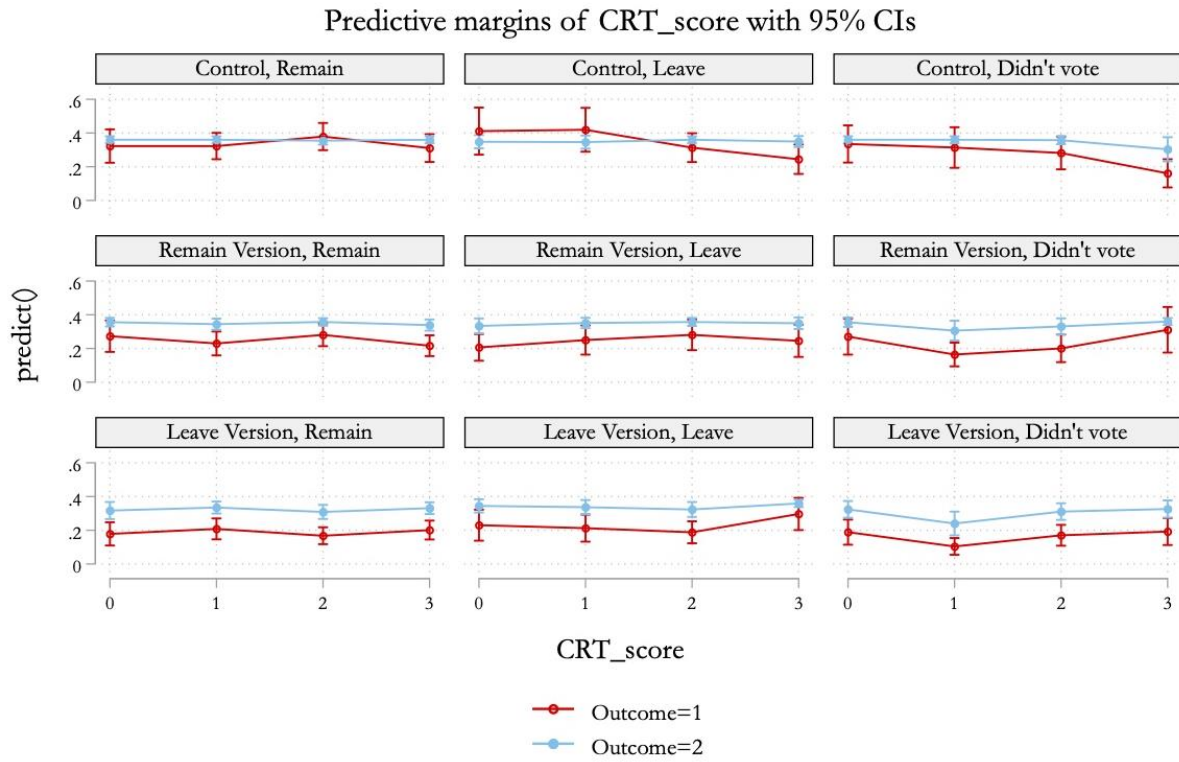
The 3-way interactions were also insignificant, although to understand them fully we should present predicted values to understand their true effect across the full range of values for the relevant variables. Figure 18-4 and Figure 18-5 thus present the probability of supporting (Figure 18-4) or opposing (Figure 18-5) the proposal by a person's CRT score, Brexit vote, and version of the experiment they received.

There is some evidence that individuals did consider their Brexit partisanship when evaluating the proposal. Remain voters were slightly less likely to strongly agree with the proposal if told it was popular in areas that voted Leave (bottom left panel, Figure 18-4 compared to top left and middle left panels, Figure 18-4), but Leave voters were not particularly less likely to strongly agree if they were told it was particularly popular in a Remain area.

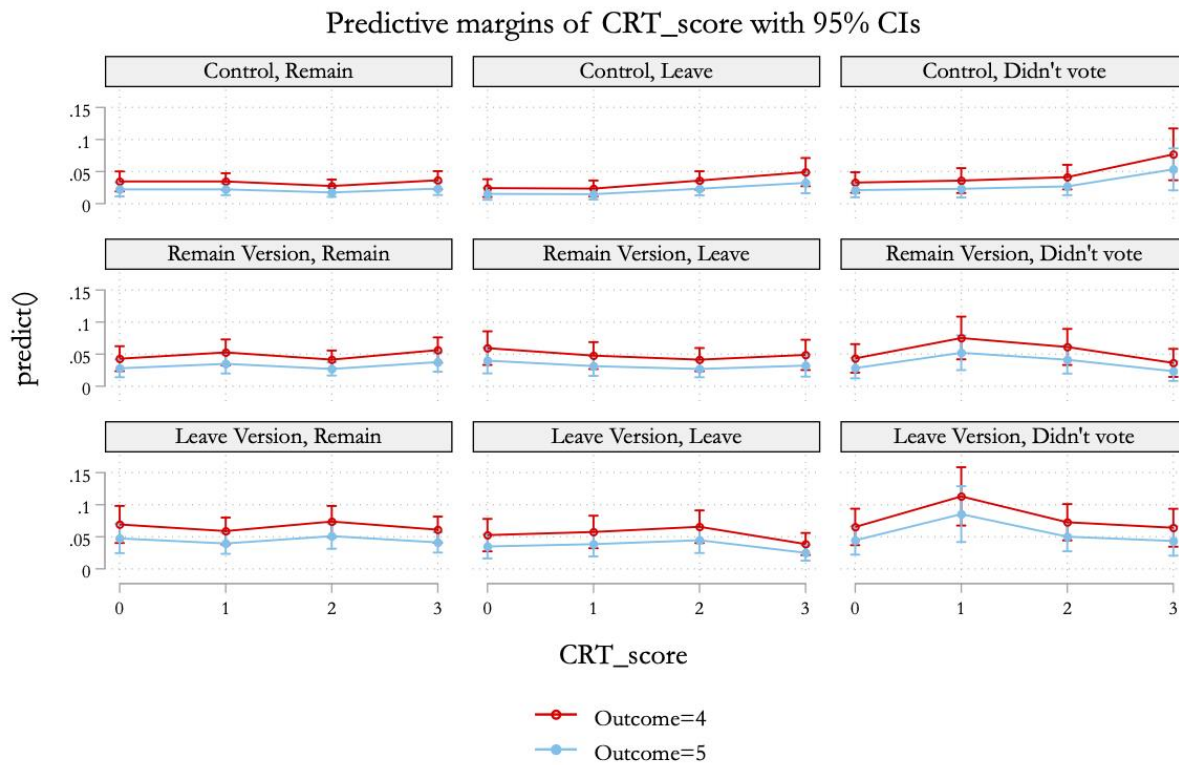
However, we do not see a strong relationship between CRT score and the extent to which an individual is more likely to agree/disagree with the proposal if they are told that it was popular in Leave/Remain areas i.e. we do not find substantial evidence that those with a high CRT score use heuristics to a greater extent than those with a lower CRT score. The only possible exception is Leave voters who received the Leave version of the experiment; it does appear that those who scored a 3 on the CRT test were slightly more likely to strongly support the proposal (middle bottom panel, Figure 18-4).



**Figure 18-4 Likelihood of supporting (outcome 2) or strongly supporting (outcome 1) the proposal. N=2354**



**Figure 18-5 Likelihood of opposing (4) or strongly opposing (5) the proposal. N=2354**



Overall, our tentative (given the nature of the data and the analysis) results indicate that CRT scores are not associated with the extent to which individuals were more or less likely to support a policy proposal when told that it had support from areas with whom they shared or did not share their Brexit identity/vote choice. This could indicate that cognitive reasoning is not associated with politically motivated reasoning; we certainly do not find much evidence to support this. However, to be confident of these results, we need to control for factors such as education, age, and gender, which are associated with CRT-scores. We also may have found evidence that Brexit identities are weakening compared to the post-election period and that the effect of PMR is only found in highly polarised contexts (the US). Certainly, we do not replicate Sumner et al.'s (2023) finding that Brexit identities lead to PMR based on data collected in 2016. We will attempt to investigate this further with additional analysis but it may be that untangling the puzzle will require further experimentation.

## **18.4 References**

- Baker, S.G., Patel, N., Gunten, C.V., Valentine, K.D. & Scherer, L.D. (2020). Interpreting politically-charged numerical information: The influence of numeracy and problem difficulty on response accuracy, *Judgment and Decision Making*, 15(2): 203-213.
- Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology*, 85(5), 808–822.
- Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2019). At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives. *Perspectives on Psychological Science*, 14(2), 273-291.
- Greene, C. M., Nash, R.A. & Murphy, G. (2021). Misremembering Brexit: Partisan Bias and Individual Predictors of False Memories for Fake News Stories among Brexit Voters. *Memory* 29(5): 587–604.

- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. Penguin Books.
- Kunda, Z. (1990). The case for motivated reasoning, *Psychol Bull*, 108(3): 480-98.
- Maguire, A., Persson, E., Västfjäll, D. & Tinghög, G. (2022). COVID-19 and Politically Motivated Reasoning, *Medical Decision Making*, 42(8):1078-1086.
- Sorace, M., & Hobolt, S. B. (2021). A tale of two peoples: motivated reasoning in the aftermath of the Brexit Vote. *Political Science Research and Methods*, 9(4): 675–692.
- Stagnaro, M.N., Tappin, B.M. & Rand, D.G. (2023) No association between numerical ability and politically motivated reasoning in a large US probability sample. *Proc Natl Acad Sci USA*, 8;120 (32): e2301491120.
- Strömbäck C, Andersson D, Västfjäll D, Tinghög G. (2021). Motivated reasoning, fast and slow, *Behavioural Public Policy*. 2021:1-16.
- Sumner, C., Scofield, J.E., Buchanan, E.M., Evans, M.R. and Shearing, M (2023). The role of personality, authoritarianism and cognition in the United Kingdom’s 2016 referendum on European Union membership, *Front. Psychol*. 14:1077354.
- Taber, C.S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769
- Taber, C.S. and Lodge, M. (2016), The Illusion of Choice in Democratic Politics: The Unconscious Impact of Motivated Political Reasoning. *Political Psychology*, 37: 61-85.
- Tappin, B. M., Pennycook, G. & Rand, D. G. (2021). Rethinking the link between cognitive sophistication and politically motivated reasoning, *Journal of Experimental Psychology: General*, 150(6): 1095-1114.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99–113.

University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, [DOI: https://doi.org/10.5255/UKDA-SN-6849-16](https://doi.org/10.5255/UKDA-SN-6849-16).

## **19. Child development measures from the ‘red book’**

*Michaela Benzeval, Institute for Social and Economic Research, University of Essex*

*John Payne, Institute for Social and Economic Research, University of Essex*

### **19.1 Introduction**

Understanding Society is a longitudinal household study, which collects data on people of all ages. This makes it a unique resource for children’s research, often having data on both parents before the child is born as well as siblings, and then annual data on the family with detailed data on the children at key milestones. For research on child development though, having early data, and in particular objective data, would be very valuable. However, directly and regularly measuring physical health is invasive and time consuming for families, and costly for studies. In wave 16 of the Innovation Panel, therefore, we investigated whether it was possible to collect such data from the NHS ‘Personal Child Health Record’ parents hold for their children, commonly known as the ‘red book’. The red book is a paper booklet where health visitors and parents record key development milestones for children under the age of 2 years. At the time of developing this experiment, some health authorities were beginning to move the red book to a digital record.

### **19.2 Methods**

The experiment is described in detail in chapter 9 (Benzeval et al, 2024, this volume). Briefly, households with dependent children were randomised into two groups. One received a request to upload photos from the red book before the interview, and the other did not. For those in the group who did not receive the letter, or those who did not upload photographs in advance, they were asked to do so during the interview. Unfortunately, the questionnaire software employed by our fieldwork agency did not allow for files to be attached, and so this had to be scripted in separate software, to which parents or interviewers were ported if they agreed to upload photos. During the photo upload process, parents were asked for the

child's name and date of birth, to check the information was linked to the correct child. When parents were not able or willing to upload a photograph, they were asked to enter the measurements themselves as part of the interview, identifying whether or not they had used the red book to do so.

After the data collection was complete, the red book photographs were returned to ISER for data entry. A wide range of red book formats were uploaded by parents, and health visitors recorded information by hand in very varied ways. Not all parents provided the pages requested. For example, some parents provided information at different ages, or pages that were only their own notes about the child's progress, or included assessment of the child's development (Including ASQ – Age and Scales Questionnaires scores) – but not the child's weight and length/height. Other times, while the photograph included measurement data, health visitors had added multiple types of measures, making it hard to identify the required information. Given this it made it very hard to create a standardised data entry process, although a range of checks were included to verify the data entered where possible.

The data was transcribed from the red book photographs and included: names and dates of birth (to confirm data were for the linked child, but not share in the public file), date of measurement (to estimate the age at measure), measurement details in the various formats provided, and other information records by the health visitor about breast feeding and smoking in the household. The inputter also recorded the kinds of other development variables available on the same photograph.

The data entered from the photographs and that entered directly within the interview, were then merged into a single file `P_REDBOOK_IP`, which is part of the main IP release for wave 16 (University of Essex, Institute for Social and Economic Research, 2024) and described below.

### **19.3 Data**

The data file `P_REDBOOK_IP` is a child level file, which only includes children if some information on the child's measurements were provided. In some cases, both parents

provided data and the data from both parents has been kept. Parents were asked to enter data for all children under 16. In total there are 209 data records entered by 132 parents. Five children were entered by both parents, so we have data for 204 unique children.

The data file includes a range of different identifiers listed in Table 19-1 below.

**Table 19-1 Identifiers included in P\_REDBOOK\_IP.**

Identifier variable name	Description	Value to analysis of children's measurement
P_HIDP	The household identifier	Link to hh file (HHSAMP) with random allocation variable (FF_REDBOOKW16); and to household data more broadly (HHRESP)
PIDP	the cross-wave identifier of the adult who provided the information	Link to adult questionnaire (INDRESP), which in this wave providing details on the responsible adult who provided the information and as well as general information on the red book. Also enables linkages to other waves.
P_CHILDPNO	The pno of the child within the household from the household grid	With P_HIDP enables you to link to other data on this child in this wave

The data file includes separate variables for measurement information entered in different ways (e.g., photographs entered pre-interview and during the interview; data entered from the red book during the interview and data entered from parent's own knowledge). Both the interview, and red book data entry, allowed for information to be entered in metric or imperial measures, although the majority were provided in metric. Where parents entered the data free text fields were provided for each unit of measurement, eg kg and grams. To make it as easy as possible for parents to enter, it was the parents' choice whether they

entered the information just in kilograms, just in grams or a used both fields. Given the multiple ways data was entered, we have also created derived variables which combine measurements into a single variable. For researchers interested in using the measurements, we recommend they uses the derived variables below. Obviously for those interested in the different ways the data were captured we still provide the original entered data. Key data included within the file are outlined in Table 19-2 below.

**Table 19-2 Key variables in P\_REDBOOK\_IP. (order appear in file)**

<b>Variable</b>	<b>Description</b>	<b>Detail of information in variable/ values</b>
P_HIDP	Household identifier (public release)	
PIDP	Cross-wave person identifier of person who provided the information (public release)	
P_CHILDPNO	Child person number	
P_RBWHICH	Have red book for child	1 Yes 2 No
P_RBSOMEWHO	Red book uploaded in pre interview group	1 Yes 2 No
P_RBWILL	Willing to upload red book for child	1 Yes 2 No
P_RBNOTSELCTD	Red book available but not uploaded for child	1 Yes 2 No
P_RBENTINT	Red book details entered manually (by parents) for this child	1 Yes 2 No
P_RBWPIMP	Weight pounds (parent inputted)	Parent entered information in free text ie they could fully enter measurement in one unit or both.
P_RBWOIMP	Weight ounces (parent inputted)	
P_RBHIMP	Height inches (parent inputted)	
P_RBWKMET	Weight kilograms (parent inputted)	Parent entered information in free text ie they could fully enter measurement in one unit or both.
P_RBWOGMET	Weight grams (parent inputted)	
P_RBHMET	Height centimetres (parent inputted)	
P_RBUSEDWEBTEL	Red book used in web survey or telephone interviews	1 Yes 2 No
P_RBUSEDIN	Red book used (parent report)	1 Yes 2 No



<b>Variable</b>	<b>Description</b>	<b>Detail of information in variable/ values</b>
P_REDBOOKTYPE	Red book uploaded	1 Take a photo 2 Upload existing image or file
P_RBIMGSOURCE	Image uploaded by interviewer or web	1 CAWI Pre interview 2 CAWI during interview 3 CAPI
P_NAMEMATCH	Red book name matches stored record	0 name not on page 1 yes
P_DOBMATCH	Red book date of birth matches stored record	1 True 2 False - Entered incorrectly into survey 3 False - Wrong date written in red book 4 False - Date of birth field present but empty 5 False - Wrong page, no date of birth field 6 Illegible
P_UPLOADEDPAGE	Red book page uploaded	1 Page requested 2 Wrong page - parents' comments only 3 Wrong page - birth page 4 Wrong page - ASQ booklet* 5 Blank page 97 other
P_BIRTHWEIGHT	Red book birthweight in kilograms	Data were entered in kilograms with 2 decimal places (eg 3.52)
P_BIRTHWEIGHTINKG	Redbook birthweight in kilograms	Kilogram units only
P_BIRTHWEIGHTING	Red book birthweight in grams	Gram unit only
P_RBVISITWEEKS	Red book visit occurred at x weeks	
P_WEIGHTINKG	Red book weight in kilograms	Kilogram units only
P_WEIGHTING	Red book weight in grams	Gram units only
P_RBLENGTH	Red book recorded length in centimetres	
P_RBHEADCIRC	Red book head circumference in centimetres	
P_BREASTFEEDING	Red book Is child still being breastfed?	1 Yes (totally or partially) 2 No
P_BREASTFEDWEEKS	Red book breastfed for x weeks	Information calculated from date breastfeeding stopped and date of birth

Variable	Description	Detail of information in variable/ values
P_SMOKEEXP	Red book smoking exposure in household	0 No one in house smokes 1 Mother smokes 2 Other householder smokes
P_RBBOOKTYPE	Type of red book	1 categorical/scored development information 2 narrative information 3 measurement card only
P_RBFOLLOWUP	Red book note that health care follow up required	1 Yes 2 No
P_RBDETSOURCE_DV	Red book details from	1 Photo or image 2 Survey entered 3 Survey entered and photo upload 4 Said would upload but didn't
P_WEIGHT_KG_DV	Derived (overall) red book weight in kilograms	Weight from all source Kilogram units only
P_WEIGHT_G_DV	Derived (overall) red book weight in grams	Weight from all source gram units only
P_HEIGHT_CM_DV	Derived (overall) red book length in centimetres	Length from all sources (cm)

\*Ages & Stages Questionnaires® (ASQ®) is used by the NHS as a reliable, accurate developmental and social-emotional screening for children between birth and age 6. It is designed to pinpoint developmental progress and catch delays in young children to identify the need for further monitoring or intervention.

In addition to the child-specific data, some further relevant variables that are not child specific are included at the adult level in the P\_INDRESP\_IP file, as listed in Table 19-3 below. These include whether the responsible adult has the red book for their children and is willing to enter data from it for all or some children. If the responsible adult is unwilling to provide data from the red book, they are asked the reasons why.

**Table 19-3 Key variables for red book module in P\_INDRESP\_IP. (order appear in file)**

Variable	Description	Values
P_HIDP	Household identifier (public release)	
PIDP	Cross-wave person identifier (public release)	
P_RBHAVE	Have red book	1 Yes 2 No

<b>Variable</b>	<b>Description</b>	<b>Values</b>
P_RBLETTER	Have received red book letter	1 Yes 2 No
P_RBDONE	Already done red book pre-interview	1 Yes 2 No 3 For some, but not all of my children
P_RBNOTALL	Willing to upload children's health information	0 None 1 Some but not all 2 All
P_RBWILL96	Willingness to upload red book - None	0 Not mentioned 1 Yes mentioned
P_RBENTINT96	Red book entry intro - None	0 Not mentioned 1 Yes mentioned
P_RBMEASURE	Measurements imperial or metric	1 Inches and pounds 2 Centimetres and kilograms
P_RBOTHMEAS	Height and weight not in red book	1 Can provide exact weight 2 Can provide a range 3 Can provide percentile 4 Unable to provide weight
P_RBNOTUSED1	Red book not used - Do not have the personal child health records	0 Not mentioned 1 Yes mentioned
P_RBNOTUSED2	Red book not used - Have the personal child health records but cannot access easily	0 Not mentioned 1 Yes mentioned
P_RBNOTUSED3	Red book not used - Reasons these are wanted are not clear	0 Not mentioned 1 Yes mentioned
P_RBNOTUSED4	Red book not used - No way to take picture	0 Not mentioned 1 Yes mentioned
P_RBNOTUSED5	Red book not used - Do not trust these will be securely handled	0 Not mentioned 1 Yes mentioned
P_RBNOTUSED97	Red book not used - Other	0 Not mentioned 1 Yes mentioned

Variable	Description	Values
P_RBNOTUSEDOTH_CODE	Red book not used - Other reason	1 Do not have the personal child health record 2 Have the personal child health record but cannot access easily 3 Reasons these are wanted are not clear 4 No way to take picture 5 Do not trust these will be securely handled 6 Did not have time 7 Not comfortable giving child's details / their personal info 8 Did not want to do it / not willing 97 Other

## 19.4 References

Benzeval, Michaela, Edith Aguirre, Tarek Al Baghal, and Lewis Mitchell. Obtaining measurement data from the 'red book'. In Vine, J. et al. *Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data*, 46–50 (2024). Understanding Society Working Paper Series 2024–11. Colchester, Essex: Institute for Social and Economic Research, University of Essex.

University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Innovation Panel, Waves 1-16, 2008-2023*. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.

## **20. Sea Hero Quest: Spatial navigation data linked to the Innovation Panel wave 16**

*Antoine Coutrot, Centre National de la Recherche Scientifique*

*Tarek Al Baghal, Institute for Social and Economic Research, University of Essex*

*Hugo Spiers, University College London*

### **20.1 Introduction**

At IP16, respondents were asked to download and play the mobile game, Sea Hero Quest. Sea Hero Quest is a smartphone- and tablet-based video game designed to measure human spatial navigation ability through gameplay (Coutrot et al., 2018). It was initially designed to aid research on dementia, although the spatial cognition measures derived from playing this game are valuable across the entire population, as it is important in several aspects of life functions.

#### **Wayfinding and path integration in spatial cognition**

Wayfinding refers to a task requiring travel through an environment to a remembered or indicated goal location. Wayfinding was chosen by the SHQ designers as it forms a common part of everyday navigation experience, and relies on a wide array of cognitive abilities, including interpretation of a map, planning a multi-stop route, memory of the route, monitoring progress along the route and updating of route plan, and transformation of a birds-eye perspective to an egocentric perspective needed for navigation.

During path integration, one integrates perceived ego motion during travel to update one's position and orientation. It is a more basic (and evolutionarily highly conserved) navigation mechanism, which typically only requires working memory processes.

Together, wayfinding and path integration capture a wide range of the abilities and processes that are required for everyday successful navigation. Sea Hero Quest captures

both of these aspects of spatial ability as part of gameplay. Spatial ability assessed by Sea Hero Quest is correlated with spatial ability in the real world (Coutrot et al., 2019).

## **20.2 Methods**

Sea Hero Quest has a portal where researchers can create projects. Researchers can choose the levels of the game (out of a possible 75) to be used in their research project, and can generate login IDs to assign to participants. For IP16, 17 levels (denoted with an 'L') were used in the game: L1, L2, L3, L4, L5, L6, L7, L8, L9, L11, L12, L23, L32, L41, L54, L74, L75. These were chosen to include measurements of levels with varying difficulties and tasks, as well as including enough levels to provide respondents the feeling of playing a full game. As a game, it is intended to be fun and less work than other app studies that IP respondents have been asked to complete in the past, and the goal was to increase participation by highlighting this fact.

Logins generated in the portal were assigned to the IP16 issued continuing sample. During the survey, respondents were invited to download and play the game, with the unique generated login for that respondent fed into the survey script for the respondent to record and use after they downloaded the game to their device.

Respondents were randomly assigned to an incentive condition, half being offered £10 or £30 to participate. The feed forward variable used:

***ff\_navapincentw16*** (1/2 each)

1 = £10 conditional incentive

2 = £30 conditional incentive

Incentives were sent to respondents playing at least one level of the game. Initial results from this experiment are reported in chapter 10 (Burton, Jäckle and Couper, 2024, this volume).

1290 respondents installed the app. The app data is stored in a JSON file, located in a GDPR-compliant server secured by a password and only accessible to the principal investigator. The JSON data were converted into metrics (discussed below) and linked back to the respondents' Understanding Society Innovation Panel identifier (PIDP) via the login provided. Two levels, 5 and 75, do not produce data as these were designed only to forward the game. Respondents completed these as part of the game, but they are not included as part of the following discussion or the released data set.

Of the 1290 respondents who installed the app, there were 2 cases where missing data was returned for all levels, so these cases were dropped from the released data set. A further 51 cases had missing data for level 1 but had data for at least one later level, so have been retained. There are multiple possible explanations for level data being missing. It could be an error in the server, or the participant quit the game before completing a level, then resumed and completed the level in a subsequent attempt. Since only the first attempt is used to avoid learning effects, the value would be recorded as missing. A further 14 cases where the app was installed have been dropped from the data set because they relate to sample members who declined to answer the main annual interview at IP16.

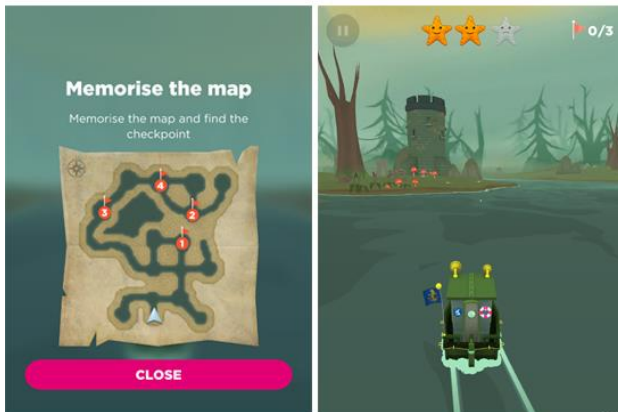
## Tasks

In **wayfinding levels** (L3, L6, L7, L8, L11, L12, L23, L32, L41), the player is initially presented with a map indicating the start location and the location of several checkpoints to find in a set order (Figure 20-1a, left). When the player has memorized the map, they hit the "close" button: the map disappears, and the player has to sail the boat as efficiently as possible to the checkpoints (Figure 20-1a, right).

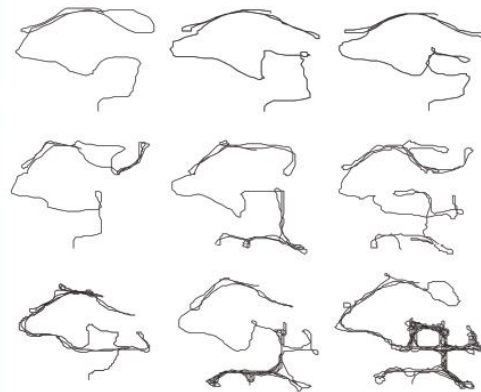
In **path integration levels** (L4, L9, L54, L74), participants navigate along a river with bends to find a flare gun and then choose which of three directions is the correct direction back to the starting point.

**Figure 20-1 Sea Hero Quest (SHQ), examples from level 42**

**a - Sea Hero Quest (SHQ), level 42**



**b - Examples of SHQ trajectories**



## Measurement

In **wayfinding levels**, a player's trajectory is sampled at 2 Hz: one (x,y) coordinate every 500 ms (Figure 20-1b). From this signal, we can compute very straightforward proxies of the player's spatial ability, such as the trajectory length (the shorter the better).

To have a single "wayfinding score", one method is to normalize the trajectory lengths (distance) within each level (z-score) and take the average of the z-score across levels. The provided data set has one such derived metric (see below). Researchers can also derive their own z-scores for each level or mean z-score across any number of levels chosen. The more levels chosen will lead to fewer cases, as there is a decrease in respondents across levels. This should be balanced against fewer levels to capture ability. Fewer levels used are more prone to getting level-specific effects (e.g. linked to the fog, or to its topology)

Duration is not used in the wayfinding scoring as it is more likely biased by gaming skills: people with less gaming experience will probably not accelerate as much and let the boat move at baseline speed. But duration is still something that can be of interest, depending on the population and hypotheses. Coutrot et al. (2018) calculated the Pearson's correlation between trajectory length and duration and found  $r = 0.75$ .



The duration and distance at a single level are both also meaningful, for instance, to look at how an effect size varies with the level of difficulty. The average z-score is used when we only need a single variable to measure spatial ability.

The first two levels are tutorial levels where no sense of direction is required, and not included in overall wayfinding scores. The outcomes of these first two levels can be used to assess the player's familiarity with video games. To have a "motor score", a z-score can also be derived from the two tutorial levels (L1, L2) in the same manner as the wayfinding levels.

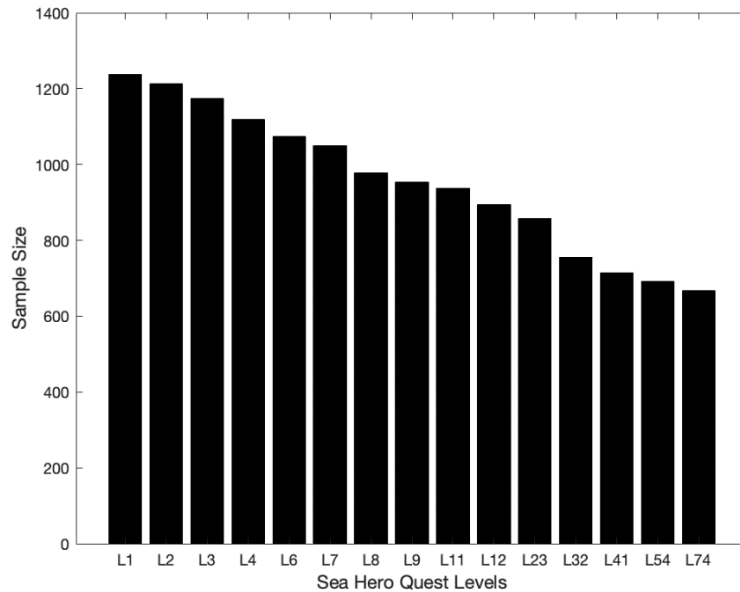
In **path integration levels**, the output is binary: the chosen direction is either correct or incorrect. In path integration levels, there is no map. The participant directly starts sailing until they find a flare gun. They are then asked to shoot the flare toward the direction of their starting position. They are given 3 choices, so the output is binary: they either chose the correct option or one of the 2 incorrect options. In Coutrot et al. (2018), wayfinding and path integration results were combined into a single metric. Using a PCA, they showed that the wayfinding data explained much more of the variance than the path integration data. They found a moderate correlation between Path Integration and Wayfinding performance (Pearson's  $r = 0.20$ ).

## **Innovation Panel**

Of the 1290 participants included in the Innovation Panel data, 96% played at least one level, and 73% played until at least level 11, see Figure 20-2.

The derived "wayfinding score" provided from the Innovation Panel data is the average of the z-score of distance in the wayfinding levels until level 11 (L3, L6, L7, L8, L11), as previous research showed that it is a good trade-off between robustness and sample size (Coutrot et al., 2022). Participants who didn't play all levels until level 11 are not included in this derived measure.

**Figure 20-2 Sample size across all levels included in the Innovation Panel**



The wayfinding score has been associated with many socio-demographic variables, including age, gender, culture, education, childhood environment, sleep (Coutrot et al., 2018, 2022a, 2022b). See (Spiers et al., 2021) for a review.

### **20.3 Data**

The data file `P_SHQ_RESULTS_IP` contains the respondent identifier (`PIDP`) and measures for each level (University of Essex, Institute for Social and Economic Research, 2024). In the training levels (L1, L2) and the wayfinding levels (L3, L6, L7, L8, L11, L12, L23, L32, L41) there are measures for distance and duration (in seconds).

Distance measures are stored in the format of `P_L#_DISTANCE`, where # represents the level played. Duration is indicated using a similar format, `P_L#_DURATION`.

Path integration levels (L4, L9, L54, L74) only have an indicator of success or not. These are dichotomised (0,1), and are stored in the variable `P_L#_DIRECTION`, where # represents the level played.

There are two derived measures included in the data file: `P_WAYFINDING_ZSCORE` and `P_TRAINING_ZSCORE`. The variable `P_WAYFINDING_ZSCORE` is the mean z-score across the wayfinding levels up to and including level 11 (L3, L6, L7, L8, L11). These were chosen as a balance between including enough levels to get a better measure of wayfinding and keeping a large enough sample of respondents who had completed these levels. Researchers can derive their own z-scores using the `P_L#_DISTANCE` variables.

The `P_TRAINING_ZSCORE` is derived as the average z-score of the distance measure in the training levels, and can be used as a “motor score”, assessing the respondent’s familiarity with video games.

There are also several metadata variables collected from the researcher portal regarding the respondents’ activity on the app. This includes the percentage of the study’s game levels completed and the date and time the last submission of game data was made. Date variables are broken into day, month and year, and time broken into the hour and minute of the day.

The variables in the data set can be summarized as shown in Table 20-1.

**Table 20-1 Summary of variables in the data set**

Variable	Value
PIPD	Respondent identifier
P_L#_DISTANCE	Distance travelled on level
P_L#_DURATION	Time spent on level in seconds.
P_L#_DIRECTION	Path integration done correctly (=1) or not (=0) in level
P_WAYFINDING_ZSCORE	Derived mean of z-scores for wayfinding levels through level 11 (L3, L6, L7, L8, L11). Lower scores indicate better wayfinding ability.
P_TRAINING_ZSCORE	Derived mean of z-scores for training levels (L1, L2). Lower scores indicate better initial ability.
P_COMPLETED	Percent of game levels completed
P_LASTSUBMISSION_D	Date of last submission, day
P_LASTSUBMISSION_M	Date of last submission, month
P_LASTSUBMISSION_Y	Date of last submission, year
P_LASTSUBMISSION_HR	Time of day of last submission, hour
P_LASTSUBMISSION_MIN	Time of day of last submission, minute

## 20.4 References

- Burton, J., Jäckle, A., and Couper, M. The effect of incentives on the willingness to participate in a gamified app study. In Vine, J. et al. *Understanding Society Innovation Panel wave 16: Results from methodological experiments and new data*, 51–61 (2024). Understanding Society Working Paper Series 2024–11. Colchester, Essex: Institute for Social and Economic Research, University of Essex.
- Coutrot, A. et al. Global Determinants of Navigation Ability. *Current Biology* **28**, 2861–2866 (2018).
- Coutrot, A. et al. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *PLOS ONE* **14**, 1–15 (2019).

Coutrot, A. *et al.* Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110 (2022a).

Coutrot, A. *et al.* Reported sleep duration reveals segmentation of the adult life-course into three phases. *Nature Communications* **13**, 7697 (2022b).

Spiers, H. J., Coutrot, A. & Hornberger, M. Explaining World-Wide Variation in Navigation Ability from Millions of People: Citizen Science Project Sea Hero Quest. *Topics in Cognitive Science* **15**, 120–138 (2021).

University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Innovation Panel, Waves 1-16, 2008-2023. [data collection]. 13th Edition. UK Data Service. SN: 6849, DOI: <https://doi.org/10.5255/UKDA-SN-6849-16>.