Understanding Society

Working Paper Series

No. 2021 – 01

February 2021

Understanding and improving data linkage consent in surveys

Annette Jäckle[1], Jonathan Burton[1], Mick P. Couper[2], Thomas F. Crossley[3] and Sandra Walzenbach[4]

[1] University of Essex

[2] University of Michigan

[3] European University Institute

[4] University of Konstanz

Understanding Society

UKRI Economic and Social Research Council

Nuffield Foundation

Table of contents

Executive summary

Motivation and Objectives

Linking survey and administrative data offers the possibility of combining the strengths, and mitigating the weaknesses, of both. Such linkage is therefore an extremely promising basis for future empirical research in social science.

For ethical and legal reasons, linking administrative data to survey responses will usually require obtaining explicit consent.[1] It is well known that not all respondents give consent. Past research on consent has generated many null and inconsistent findings. A weakness of the existing literature is that little effort has been made to understand the cognitive processes of how respondents make the decision whether or not to consent.

The overall aim of this project was to improve our understanding about how to pursue the twin goals of maximizing consent and ensuring that consent is genuinely informed. The ultimate objective is to strengthen the data infrastructure for social science and policy research in the UK. Specific aims were:

1. To understand how respondents process requests for data linkage: which factors influence their understanding of data linkage, which factors influence their decision to consent, and to open the black box of consent decisions to begin to understand *how* respondents make the decision.
2. To develop and test methods of maximising consent in web surveys, by understanding why web respondents are less likely to give consent than face-to-face respondents.
3. To develop and test methods of maximising consent with requests for linkage to multiple data sets, by understanding how respondents process multiple requests.
4. As a by-product of testing hypotheses about the previous points, to test the effects of different approaches to wording consent questions on informed consent.

[1] In the UK, the Digital Economy Act gives the Office for National Statistics a legal right to access and link data held by a range of bodies to produce official statistics. In addition, European GDPR legislation means that if a survey is conducted in the 'public interest', then consent is not required for legal reasons, although it may still be required for ethical reasons.

Methods and Data

Our findings described below are based on a series of experiments conducted in four surveys using two different studies: The *Understanding Society* Innovation Panel (IP) and the PopulusLive online access panel (AP). The Innovation Panel is part of *Understanding Society*: the UK Household Longitudinal Study. It is a probability sample of households in Great Britain used for methodological testing, with a design that mirrors that of the main *Understanding Society* survey. The Innovation Panel survey was conducted in wave 11, fielded in 2018. Since the Innovation Panel sample size (around 2,900 respondents) constrained the number of experimental treatment groups we could implement, we fielded a parallel survey with additional experiments, using a different sample. PopulusLive is a non-probability online panel with around 130,000 active sample members, who are recruited through web advertising, word of mouth, and database partners. We used age, gender and education quotas to match the sample composition of the Innovation Panel.

A total of nine experiments were conducted across the two sample sources. Experiments 1 to 5 all used variations of a single consent question, about linkage to tax data (held by HM Revenue and Customs, HMRC). Experiments 6 and 7 also used single consent questions, but respondents were either assigned to questions on tax or health data (held by the National Health Service, NHS) linkage. Experiments 8 and 9 used five different data linkage requests: tax data (held by HMRC), health data (held by the NHS), education data (held by the Department for Education in England, DfE, and equivalent departments in Scotland and Wales), household energy data (held the Department for Business, Energy and Industrial Strategy, BEIS), and benefit and pensions data (held by the Department for Work and Pensions, DWP).

The experiments, and the survey(s) on which they were conducted, are briefly summarized here:

1. Easy vs. standard wording of consent request (IP and AP). Half the respondents were allocated to the 'standard' question wording, used previously in *Understanding Society*. The balance was allocated to an 'easy' version, where the text was rewritten to reduce reading difficulty and to provide all essential information about the linkage in the question text rather than an additional information leaflet.
2. Early vs. late placement of consent question (IP). Half the respondents were asked for consent early in the interview, the other half were asked at the end.
3. Web vs. face-to-face interview (IP). This experiment exploits the random assignment of IP cases to explore mode effects on consent.

4. Default question wording (AP). Experiment 4 tested a default approach to giving consent, asking respondents to *"Press 'next' to continue"* or explicitly opt out, versus the standard opt-in consent procedure.
5. Additional information question wording (AP). This experiment tested the effect of offering additional information, with a version that added a third response option ("*I need more information before making a decision"*) to the standard 'yes' or no' options.
6. Data linkage domain (AP). Half the respondents were assigned to a question asking for consent to link to HMRC data; the other half were asked for linkage to NHS data.
7. Trust priming (AP).This experiment was crossed with the data linkage domain experiment, and focused on the effect of priming trust on consent. Half the sample saw an additional statement: "*HMRC / The NHS is a trusted data holder"* on an introductory screen prior to the consent question. This was followed by an icon symbolizing data security: a shield and lock symbol with the heading *"Trust".* The balance was not shown the additional statement or icon.
8. Format of multiple consents (AP). For one group, the five consent questions were each presented on a separate page, with respondents consenting to each in turn. For the second group the questions were all presented on one page; however, the respondent still had to answer each consent question individually. For the third group all five data requests were presented on a single page and the respondent answered a single yes/no question, whether they consented to all the linkages or not.
9. Order of multiple consents (AP). One version asked the five consent questions in ascending order of sensitivity of the request (based on previous data), with NHS asked first. The other version reversed the order, with consent to linkage to HMRC data asked first.

For all of the experiments described above, we examined the rates of consent. We also tested comprehension of the consent request, using a series of knowledge questions about the consent process. We also measured subjective understanding, to get a sense of how much respondents felt they understood about the request. Finally, we also ascertained subjective confidence in the decision they had made.

In additional to the experiments, we used digital audio-recordings of the IP11 face-to-face interviews (recorded with respondents' permission) to explore how interviewers communicate the consent request to respondents, whether and how they provide additional information or attempt to persuade respondents to consent, and whether respondents raise questions when asked for consent to data linkage.

Key Findings

Correlates of consent:

(1) Respondents who have better understanding of the data linkage request (as measured by a set of knowledge questions) are also more likely to consent.
(2) As in previous studies, we find no socio-demographic characteristics that consistently predict consent in all samples. The only consistent predictors are positive attitudes towards data sharing, trust in HMRC, and knowledge of what data HMRC have.
(3) Respondents are less likely to consent to data linkage if the wording of the request is difficult and the question is asked late in the questionnaire. Position has no effect on consent if the wording is easy; wording has no effect on consent if the position is early.
(4) Priming respondents to think about trust in the organisations involved in the data linkage increases consent.
(5) The only socio-demographic characteristic that consistently predicts objective understanding of the linkage request is education. Understanding is positively associated with the number of online data sharing behaviours (e.g., posting text or images on social media, downloading apps, online purchases or banking) and with trust in HMRC.
(6) Easy wording of the consent question increases objective understanding of the linkage request. Position of the consent question in the questionnaire has no effect on understanding.

The consent decision process:

(7) Respondents decide about the consent request in different ways: some use more reflective decision-making strategies, others use less reflective strategies.
(8) Different decision processes are associated with very different levels of consent, comprehension, and confidence in the consent decision.
(9) Placing the consent request earlier in the survey increases the probability of the respondent using a reflective decision-making process.

Effects of mode of data collection on consent:

(10) As in previous studies, respondents are less likely to consent online than with an interviewer.
(11) Web respondents have lower levels of understanding than face-to-face respondents.
(12) There is no difference by mode in respondents' confidence in their decisions.
(13) Web respondents report higher levels of concern about data security than face-to-face respondents.

(14) Web respondents are less likely to use reflective strategies to make their decision than face-to-face respondents, and instead more likely to make habit-based decisions.
(15) Easier wording of the consent request does not reduce mode effects on rates of consent.
(16) Respondents rarely ask questions and interviewers rarely provide additional information.

Multiple consent requests:

(17) The format in which a sequence of consent requests is asked does not seem to matter.
(18) The order of multiple consent requests affects consent rates, but not in a consistent way.
(19) Objective knowledge, subjective understanding and subjective confidence in the decision do not differ much by order and format of sequential consent requests.
(20) The order effects of multiple consent requests from Study 1 do not replicate in Study 2.

Conclusions and Recommendation

This series of studies has shed light on some of the processes underlying the consent process and offered a theoretical framework for better understanding how the consent decision is made. The different decision processes employed by survey respondents are associated with different levels of consent, comprehension, and confidence in the consent decision. Generally, respondents reach a consent decision relatively quickly. Given this, simply providing more information on the consent process is unlikely to be effective. Rather, wording consent requests in an easy-to-read format and emphasising trust in the organisations involved will likely increase rates of consent without compromising understanding of the request or confidence in the decision.

This research has advanced our understanding on how the decision to consent to administrative data linkages is made. It points to the importance of understanding how respondents process the request for consent in different ways, suggesting that targeting different strategies based on respondents' decision-making preferences may be effective at increasing informed consent. Our work also points to the importance of focusing not only on the outcome of the request (i.e., maximizing consent rates) but also on understanding how informed the consent is, measured both objectively and subjectively. However, more work remains to achieve the goal of maximizing informed consent to administrative record linkage in surveys, especially those administered online.

Understanding and improving data linkage consent in surveys

Annette Jäckle (University of Essex), Jonathan Burton (University of Essex), Mick P. Couper (University of Michigan), Thomas F. Crossley (European University Institute) and Sandra Walzenbach (University of Konstanz)

Abstract: This report summarizes findings from a project that aimed to improve our understanding about how to pursue the twin goals of maximizing consent to data linkage and ensuring that consent is genuinely informed. Using experimental data from the *Understanding Society* Innovation Panel and the PopulusLive online access panel, we examined (1) how respondents process requests for data linkage, (2) why web respondents are less likely to give consent than face-to-face respondents, (3) how respondents process multiple requests for consent, and (4) what effects different approaches to wording consent questions have on informed consent.

Keywords: Informed consent, consent rates, decision process, administrative data linkage, survey mode, mode effects, multiple consent requests, question wording

Acknowledgements: This research was funded by the Nuffield Foundation with co-funding from the ESRC (OSP/43279). Data collection on the Innovation Panel was funded by the ESRC grant for *Understanding Society*: The UK Household Longitudinal Study, Waves 9-11 (ES/N00812X/1).

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

Understanding Society is an initiative funded by the Economic and Social Research Council 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 Kantar Public. The research data are distributed by the UK Data Service.

We are also grateful for the feedback and suggestions received from the project advisory board.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, University of Essex, aejack@essex.ac.uk.

1. Introduction

One of the most promising avenues for empirical social science research involves linking administrative or process generated data with survey data. Both administrative data (whether held by government or private entities) and survey data are useful on their own, but they have complementary strengths, so that "integrated" data that combines both is particularly attractive (Groves 2011; Davis-Kean et al. 2017; Benzeval et al. 2020). Surveys can be used to "fill the gaps" in administrative data, and the other way around. Sometimes the gaps will be specific types of information (e.g., administrative data do not contain information on expectations or subjective wellbeing), and sometimes the gaps will be in population coverage (e.g., surveys have poor coverage of very high-income households, while tax data may lack information on non-filers). In some cases, administrative data may provide a suitable survey frame to allow inference to the general population (especially in the UK where there is not an appropriate individual identifier, or register, to provide a frame).

Depending on the legal framework, it is sometimes legally necessary to obtain informed consent from survey respondents, before linking their data to administrative or other process-generated data. In the UK, the Digital Economy Act gives the Office for National Statistics a legal right to access and link data held by a range of bodies to produce official statistics. In addition, European GDPR legislation means that if a survey is conducted in the 'public interest', then consent is not required for legal reasons. However, consent might still be necessary for ethical or practical reasons. Respondents' participation in most surveys depends on their good will and trust in the survey organisation. If their responses can be onward linked without specific consent, this could attenuate participation. This concern may be particularly sharp in longitudinal surveys, where participant retention is critical. So, for a varying mix of legal, ethical and practical reasons, survey organisations in the UK and elsewhere continue to seek consent to data linkage for the foreseeable future.

Previous research has shown that not everyone gives consent, and that the probability of consent varies by topic matter (Sakshaug et al. 2012), between interviewers (Korbmacher and Schroeder 2013; Sakshaug, Tutz and Kreuter 2013), and even within respondents over time (Weir, Faul and Ofstedal 2014). This evidence suggests that people do not have strong fixed views on consent and that the decision to consent can be influenced. However, studies examining what influences the decision to consent have had limited success, with many null findings and inconsistent findings across studies (see Section 2).

Many large-scale surveys in the UK (and abroad) are moving to mixed mode data collection where some sample members complete the survey by web. Evidence from earlier waves of the

Understanding Society Innovation Panel suggests that respondents are 30 percentage points less likely to give consent by web than in a face-to-face interview (Jäckle et al in press). A pilot study for the Next Steps cohort study reached similar conclusions (Calderwood 2016; Thornby et al. 2018). Generally, not much is known about why consent rates are lower in web, or how to best obtain informed consent from web respondents.

Finally, most surveys aiming to link to administrative data ask for consent to more than one linkage at a time. Previous research on consent has however focused on single consent questions and not much is known about how respondents process multiple consent requests.

In our view, a weakness of the existing literature is that little effort has been made to understand the cognitive processes of how respondents make the decision whether or not to consent.

The overall aim of this project was to improve our understanding about how to pursue the twin goals of maximizing consent and ensuring that consent is genuinely informed. The ultimate objective is to strengthen the data infrastructure for social science and policy research in the UK. Specific aims were:

1. To understand how respondents process requests for data linkage: which factors influence their understanding of data linkage, which factors influence their decision to consent, and to open the black box of consent decisions to begin to understand *how* respondents make the decision.
2. To develop and test methods of maximising consent in web surveys, by understanding why web respondents are less likely to give consent than face-to-face respondents.
3. To develop and test methods of maximising consent with requests for linkage to multiple data sets, by understanding how respondents process multiple requests.
4. As a by-product of testing hypotheses about the previous points, to test the effects of different approaches to wording consent questions on informed consent.

This project complements an initiative by the Nuffield Foundation, CLOSER, ESRC and RSS to overcome the legal and practical barriers in using administrative data for survey purposes (The Nuffield Foundation 2016). While that initiative focuses on issues of data linkage and access to administrative data, an often-necessary precursor is that survey respondents give informed consent to such linkage.

The rest of this report is structured as follows. Section 2 lays out the background to our work by reviewing what was previously known about consent to data linkage. Most of this draws on the published literature on consent to data linkage, but 2.6 summarizes qualitative research that we had undertaken previously to inform the research design for this project, and 2.7 draws lessons from the

broader literature on decision making. The latter is particularly relevant to our efforts to understand *how* survey respondents decide on how to answer a consent request. Section 3 then gives an overview of the data and methods employed in this project. Section 4 is the core of the report and summarizes our key findings. These are divided into subsections on the effects of question wording on consent (4.1), analyses of how respondents decide whether to consent (4.2), understanding the mode effect in consent rates (4.3), and exploring the best way to ask for consent to multiple data linkages (4.4). Finally, Section 5 summarizes the implications of this project for best practice and lays out next steps for future research.

2. Background

In this section we review what is known from previous literature on consent to data linkage in surveys, and related literature of relevance to the data linkage consent process.

2.1 Correlates of consent

Much of the survey literature on administrative data linkage consent has focused on correlates and covariates of consent. In many cases this is based on secondary analysis of survey data where consent to linkage was asked of respondents. Previous research has shown that the probability of consent varies by topic matter (Sakshaug et al. 2012), between interviewers (Korbmacher and Schroeder 2013; Mostafa 2016; Sakshaug, Tutz and Kreuter 2013), and even within respondents over time (Mostafa and Wiggins 2018; Weir, Faul and Ofstedal 2014).

Several studies have found systematic differences between consenters and non-consenters based on sociodemographic characteristics (see Kho et al. 2009; Dunn et al. 2004; Sakshaug et al. 2012); however, there is little consistency across studies. For example, some studies find that older people are more likely to consent (Bryant et al. 2006; Dunn et al. 2004), while other studies find that younger people are more likely to consent (Huang et al. 2007; Yawn et al. 1998), and yet some studies find no age effect (Buckley et al. 2007; Harris et al. 2005). Gender and income exhibit similar inconsistent relationships with the likelihood of consent, being significantly related to consent in some studies (Huang et al. 2007; Olson 1999) but not others (Al-Shahi, Vousden and Warlow 2005; Harris et al. 2005). On the other hand, racial and ethnic minorities are generally less likely to consent (Mostafa 2016; Sakshaug et al., 2012; Sala, Burton and Knies 2012). In general, demographic variables are only mildly associated with the propensity to consent.

A number of studies have gone beyond sociodemographic variables to look at respondent attitudes and behaviour that may be related to consent. Sakshaug et al. (2012) found that privacy and

confidentiality attitudes were strongly related to consent (see also Bates 2005; Sala, Burton and Knies 2012). Civic-mindedness was positively associated with consent (Sala, Burton and Knies 2012). Resistant respondents, that is, those who only completed the survey after multiple contacts or reminders, had lower consent rates (Sakshaug et al. 2012), as had those who had missed a survey wave in the past (Mostafa 2016). In contrast, Sala, Burton and Knies (2012) found a negative relationship between consent and length of time in the panel. Sakshaug et al. (2012) found that inattentive respondents (based on interviewer ratings) tend to provide linkage consent at a higher rate, but this finding has not been replicated.

While there is evidence that interviewers have an important effect on consent rates, efforts to explain these effects have also yielded mixed results (see Fulton 2012). One consistent finding is that there is little empirical support for the hypothesis that interviewer attitudes and personality matter (Sala, Burton and Knies 2012; Sala, Knies and Burton 2014; Sakshaug, Tutz and Kreuter 2013). That is, the mechanisms by which interviewers influence the decision to consent (or not) are still unknown.

In summary, the literature on correlates of consent suggests that observable respondent and interviewer characteristics do not explain much of the variation on consent rates. Respondent attitudes and behaviour have bigger effects, suggesting that any attempts to understand the consent process must go beyond the collection of observable characteristics. Sakshaug et al. (2012) posited a number of mechanisms of consent, but few of these have been empirically tested. To our knowledge, no studies have explored the process by which survey respondents make consent decisions.

2.2 Question wording experiments

Several studies have experimentally manipulated various features of the consent request to explore effects on consent rates.

Pascale (2011) reports on a study that varied whether the request mentioned accuracy of the data, reduction of costs, or time saving reasons. The first two are benefits to the agency, while the last was framed as a benefit to the respondent. None of these had any effect on consent rates. More recent experiments have varied the framing of the request, with gain framing mentioning the benefits from consent and loss framing mentioning the drawbacks of non-consent. These studies have yielded mixed results. Kreuter, Sakshaug and Tourangeau (2016) found that loss framing (emphasising that not linking will reduce the value of the respondent's survey data) increased consent compared to gain framing which emphasised the value of linkage. However, Sakshaug, Wolter and Kreuter (2015) found that gain framing was more effective than loss framing. Similarly,

gain framing (expressed as a time saving for the respondent) yielded higher consent rates than a neutral framing in one study (Sakshaug and Kreuter 2014) but not in another (Sakshaug, Tutz and Kreuter 2013). Sakshaug and colleagues (2019) found that the effect of framing (gain vs. loss) was evident only in one mode (a web survey, but not in a telephone survey), where loss framing yielded a higher consent rate than gain framing, but only when the consent request came at the end of the survey.

Other studies have varied when the request for record linkage consent is made. Both Eisnecker and Kroh (2016) and Sala, Knies and Burton (2014) found no effect of asking for linkage in an earlier versus later wave of a longitudinal study. Within a survey wave, asking for consent after a module of questions related to the content of the data to be linked increases consent compared to asking at the end of the questionnaire (Sala, Knies and Burton 2014), and asking it at the beginning of the survey rather than the end has a positive effect (Sakshaug et al. 2019).

Taken together, these studies suggest relatively modest effect of wording manipulations on consent to record linkage. They also point to difficulty replicating key findings across different samples and settings, including both mode and the type of linkage request being made. This, together with the findings reviewed earlier, suggests that the decision to consent to administrative record linkage is not a fixed or stable decision, but varies with features of the request.

Finally, almost none of the studies reviewed above examine outcomes beyond the consent decision itself. In one exception, Das and Couper (2014) tested understanding of what the respondents were asked to consent to, in a study of an opt-out consent option. Edwards and Biddle (in press) have also tested comprehension of the consent request. Both studies point to relatively poor comprehension of the consent process. No studies (to our knowledge) of administrative record linkage consent in surveys have examined subjective confidence in the consent decision. Both comprehension and confidence are more commonly used as outcomes in the psychological and medical literature (see Section 2.5).

2.3 Effect of mode of data collection on consent

A clear finding in the survey literature is that consent to data linkage is significantly and substantially lower in self-administered modes than when an interviewer is making the request. Jäckle and colleagues (in press) report a consent rate of 67% for those interviewed face-to-face, compared with 48% for those who responded by web in Wave 9 of the *Understanding Society* Innovation Panel. These differences persist when looking at assigned mode or using covariate adjustment or instrumental variable analysis to account for selection differences into mode. Sakshaug and

colleagues (2017) reported a "strikingly lower" linkage consent rate in the self-administered mode (mail/web; 54%) than in the interviewer-administered mode (face-to-face; 94%) in a study with random assignment to mode in Germany. Thornby and colleagues (2018) also found differences in consent rates by mode, ranging from 89% for face-to-face, 90% for telephone, and 69% for web, in the Next Steps Age 25 Survey which employed a sequential mixed-mode design. Similar mode differences are reported by Al Baghal and colleagues (2019) for social media linkage requests.

Together these results suggest that the increased use of the web in mixed-mode surveys raises significant challenges for administrative record linkage consent. Despite these striking findings, generally not much is known about why consent rates are lower in web, or how to obtain informed consent from web respondents.

2.4 Asking for multiple consents to data linkage

So far, methodological research on data linkage has largely focused on single consent questions. In analyses of multiple consent requests, both Jenkins and colleagues (2006) and Mostafa (2016) found variation in correlates of consent across different domains within the same survey. They also both found some consistency in the consent decision across domains, suggesting some latent "consent propensity". Thornby and colleagues (2018) report on qualitative interviews and field experience asking for multiple consents in the Next Steps Age 25 Study. But research on how best to ask a series of consent requests is still lacking.

One recent exception is a survey experiment on order effects by Weiß and colleagues (2019), who hypothesized a possible fatigue or ceiling effect: "As respondents do not know how many data linkage requests they will see, they might be willing to consent to the first ones, and then reach a critical point where they are not willing to share more information. Besides such a ceiling effect, every question is an additional intrusion into privacy, so that also the contextual sensitivity of the request might increase with each question" (p. 2). They found higher consent rates to the first consent request of a sequence, irrespective of the consent domain.

Given the rising demand for administrative record linkages, surveys are increasingly asking for consent to multiple domains. This suggests a need for further research in this area.

2.5 Consent in other fields of research

While there is a large literature on informed consent in other fields, especially medical/health research and psychology, research in these fields is often based on volunteers, so the issue of maximising consent rates is less of a concern, and there is a greater focus on understanding of the

consent request. Without attempting an exhaustive review of this extensive literature, we note a few exemplary studies here.

A common finding is that patients have generally poor knowledge about medical decisions (see Fagerlin et al. 2010). Similarly, Stunkel and colleagues (2010) noted that "research volunteers often do not understand critical aspects of the research in which they are participating". Given this, there are many studies focusing on different ways to present complex information about risks and benefits of clinical procedures to improve patient understanding. For example, Tait et al. (2013) showed that understanding the details of a hypothetical medical intervention was significantly increased when information documents were written at easier reading level or included graphics. Stunkel et al. (2010) experimented with standard versus concise consent forms for a drug trial and found similar rates of comprehension.

Even in consequential medical decisions, participants do not seem to invest significant effort in reading and understanding consent forms. For example, McNutt and colleagues (2008) had observers record how long people spent reading the paper forms and found that no subject spent even half the time expected based on average reading speeds. Similar findings are observed in the psychological literature, where the decision to participate in an experiment is less consequential. For example, Perrault and McCullock (2019) found that none of the subjects accepted the offer of additional information. Ripley and colleagues (2018) similarly concluded that participants "overwhelmingly chose not to read the informed consent forms before deciding to participate in the study." Similar evidence is reported by Ghandour, Yasmine and El-Kak (2013).

Given the relatively poor performance on objective knowledge tests, researchers argue for the measurement of both specific knowledge and gist knowledge (or general impressions; see Hawley et al. 2008). In this vein, Sepucha and colleagues (2010) examined the relationship between objective knowledge and subjective feelings of being informed. They found that while understanding of the medical decisions they made was relatively poor, respondents generally felt informed and were confident in their decisions. Trust in the doctor was associated with feeling informed, suggesting a "rational delegation of consent," that is, if respondents trust the organisations involved, they might feel they do not need to know the details of the request in order to consent.

This research suggests that requiring respondents to fully process and understand all of the information provided before consenting is setting a very high bar. Most participants are comfortable with the decisions they made, even without being "fully informed" in an objective sense. Providing more information may in fact be counter productive.

2.6 Qualitative research on consent

Another source of input into the research design for this project is the qualitative research we commissioned with prior funding from the ESRC (see Beninger et al. 2017). A total of 25 in-depth interviews were conducted with *Understanding Society* respondents to explore how people decide whether or not to give consent to link government administrative data to their responses in the survey. Discussions focused on two specific consent questions: consent to link to benefits and pensions data held by the Department for Work and Pensions (DWP) and health data held by the National Health Service (NHS).

Several key findings emerged from the in-depth interviews. There was a lot of confusion about what data linkage is: which organisation shares which personal data about the respondent, how data would be shared, and for what purpose. Whether or not respondents understood the consent request was affected by their literacy level, cognitive ability, prior knowledge about data sharing, specific words and phrases in the question and the supporting information booklet. Different factors were found to influence the willingness to consent to data linkage: subconscious factors including personality traits (e.g. open, trusting vs. private, suspicious); rational factors including the belief that they have nothing to hide, or nothing the survey organisation does not already know, vs. concern about potential personal repercussions or limited understanding of the data linkage request; social factors including attitudes to data sharing (everyone shares everything already vs. can never be too cautious), attitudes to data protection (confidence in security vs. scepticism), data sharing behaviours (use of social media), relevance of request to individual, experience with the organisation holding respondent records; and environmental factors including trust in the survey organisation, having time to think about the request, media coverage of fraud and data leaks, personal experience, time of day when respondents are tired, completing a survey online, and the presence of a researcher. Taken together the different factors distilled into two types of costs of consenting to data linkage for respondents: loss of privacy (a dislike for sharing much information about themselves) and risks to data security (concern about the risks of sharing their information). The perceived costs were however sometimes overridden by other factors: for example, the length of time in the panel (with no examples of misuse of their information, and evidence of data being used to further knowledge), trust in the Institute for Social and Economic Research (ISER) as deliverer of the survey, participation of their family members (they don't have an issue, so why should I). Benefits of data linkage were thought to accrue only to third parties (academic and policy researchers, and society in general). Respondents had heightened security concerns about web surveys and indicated that interviewers were key in building trust and answering questions about data linkage. When shown a request for linkage to multiple data sources respondents were generally

overwhelmed by the amount of information. They were unfamiliar with some of the organisations, uncertain about what information these organisations hold about them, unclear about the rationale for linking to data from so many organisations, and generally had great concern around financial organisations.

2.7 Literature on decision making

Finally, there are a number of overlapping literatures on decision making, judgement under uncertainty or persuasion. A common view is that the process by which people make decisions can be of two broad types, variously called systematic versus heuristic processing (Chaiken 1980), central versus peripheral processing (Petty and Caccioppo 1986), reflective versus impulsive processing (Strack and Deutsch 2004), or system 2 versus system 1 processing respectively (Kahneman 2011). The first path or process is viewed as rational, deliberate, effortful, conscious, and often reliable. The second path is viewed as unconscious, automatic, or "fast and frugal" (Gigerenzer and Goldstein 1996).

A key notion is that people are "cognitive misers" (see Corcoran and Mussweiler 2010), striving to process information efficiently and to make decisions without consuming too many cognitive resources, even if doing so may potentially compromise the accuracy of the results. That is, people often rely on heuristics to reduce complex cognitive tasks to more simple operations. A common view, popularised by the work of Tversky and Kahneman (1974; see also Kahneman 2011) is that heuristic decisions are error-prone. In contrast, Gigerenzer and Gaissmaier (2011) take the view that heuristics can be more accurate than more complex strategies even though they process less information. Gigerenzer and Gaissmaier (2011, p. 454) define a heuristic as "a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods."

Where does the request for consent to administrative record linkage fit into this literature? It potentially has more significant consequences for participants than the hypothetical choice experiments often used to test heuristic strategies. But the stakes are lower than, say, in the field of medical or financial decision-making. The decision is made on the basis of limited information (e.g., the risks of disclosure are unknown). While consent materials are often written to provide as much information as possible about the linkage process, assuming a careful evaluation of the pros and cons, consent decisions seem to be made relatively quickly and change over time. Further, the decision is an unbalanced binary choice, there being few tangible benefits of a "yes" decision and negligible consequences of a "no" decision for the individual. Contrast this, for example, with a choice between surgery and medication in medical decisions.

There are two potential implications of this brief review. The first is that the process by which survey respondents make decisions to consent (or not) to administrative data linkage needs to be better understood. If they mostly use heuristic processes, which heuristics are important? The second is that if respondents are making "fast and frugal" decisions about data linkage, then efforts targeted at increasing a reflective evaluation of pros and cons may be less effective than strategies designed to facilitate the quick decisions that they are likely to make.

3. Methodology

To begin to 'open the box' and examine how respondents make the decision whether or not to consent to data linkage, we implemented a series of question wording and survey design experiments and examined the resulting data using quantitative analysis methods. In these surveys we collected background information about the respondent, asked them for consent to link data, and then asked a series of questions about how they had made that decision, their understanding of the request, etc.

3.1 Samples

Overall, we implemented four surveys using two different studies: the *Understanding Society* Innovation Panel and the PopulusLive online access panel.

The Innovation Panel is a probability sample of households in Great Britain that is used for methodological testing and experimentation. It is part of *Understanding Society:* The UK Household Longitudinal Study and its design mirrors that of the main panel. The Innovation Panel was first fielded in 2008 with an achieved sample of 1,500 households. Interviews are sought with all household members aged 16+ once a year. To maintain the sample size, refreshment samples are added every few years. We implemented our study in wave 11, which was fielded in May to October 2018 by Kantar Public and NatCen Social Research (University of Essex. Institute for Social and Economic Research 2019). In the text below we refer to this survey as IP11.[2] Where we present results separately for face-to-face and web respondents, we refer to IP11f and IP11w, respectively. The achieved sample size for the IP11 survey was 2,896 respondents. Since the sample size

[2] For more information on the design and implementation of the Innovation Panel, see the User Guide at https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide.

constrained the number of experimental treatment groups we could implement, we fielded a parallel survey with additional experiments, using a different sample.

The second study was the PopulusLive online access panel. This is a non-probability online panel with around 130,000 active sample members, who are recruited through web advertising, word of mouth, and database partners. To enable some comparison with the Innovation Panel sample, quotas based on age, gender and education were set, that matched the characteristics of the Innovation Panel sample. Once the target number of respondents for a quota was reached, the survey was closed to further respondents with that characteristic. Two samples were selected in this way. The first was surveyed in May 2018 and a sub-set was surveyed again in May 2019. In the following text we refer to the surveys from this two-wave panel as AP1.1 and AP1.2, where AP refers to the study (access panel), 1. refers to the first sample drawn and .1/.2 refers to the first or second wave. The second sample was selected in December 2019 and surveyed only once. We refer to this survey as AP2. The implementation of these surveys was led by NatCen Social Research, in collaboration with the PopulusLive panel. The achieved sample sizes for these surveys were 5,684 respondents in AP1.1, 2,064 respondents in AP1.2, and 3,850 respondents in AP2.

3.2 Questionnaires

We implemented the same questionnaires in all surveys, with some differences: the Innovation Panel questionnaire included additional content not related to this project, we fielded different experiments in the different surveys (see Table 3.3.1 below), there are a couple of questions that we introduced in the later surveys, and one question where we introduced an additional response category. Other than that, the question wording and response options were the same in all surveys.

The questionnaires included a module of background characteristics, including attitudinal questions about privacy and data security concerns. This module preceded the consent question module, which was followed by a series of questions about how well the respondent felt they had understood the data linkage request (subjective understanding), how confident they were in their decision, how they had processed the question, and a series of true/false knowledge questions about the data linkage to measure objective understanding. This module also included a question about how much effort the respondent had put into making the consent decision (AP1.2 and AP2 only), and a question asking how much of a role different aspects played when the respondent was deciding whether or not to consent (AP2 only). This module also included questions about how sensitive respondents think data held by different government departments are and how much they trust different organisations involved in the linkages.

The main questions used for the analyses in this report are documented in the Appendix. The full questionnaires are available from https://www.iser.essex.ac.uk/research/projects/understanding-and-improving-data-linkage-consent-in-surveys.

In addition to the questionnaire data, we also captured paradata, which are generated by the computerized questionnaire scripts. These include time stamps, used to measure how long a respondent spent on a given question and, for online respondents, whether they clicked on links to additional information about the data linkage.

In the face-to-face interviews we in addition asked respondents for permission to record parts of the interview. The audio-recording was set to automatically turn on at the start of the consent question and off at the start of the following question. Interviewers were alerted to the recording by a signal on their laptop, which switched itself on and off with the recording.

3.3 Experiments

Table 3.3.1 summarizes the experiments that we implemented on the different surveys. Experiments 1 to 5 all used variations of a single consent question, about linkage to tax data. Experiments 6 and 7 also used single consent questions, but respondents were either allocated to tax or health data linkage. Experiments 8 and 9 used five different data linkage requests: tax data (held by HM Revenue and Customs, HMRC), health data (held by the National Health Service, NHS), education data (held by the Department for Education in England, DfE, and equivalent departments in Scotland and Wales, EDUC), household energy data (held by the Department for Business, Energy and Industrial Strategy, BEIS), and benefit and pensions data (held by the Department for Work and Pensions, DWP).

Table 3.3.1: Summary of experiments by survey

No.	Experiment	IP11	AP1.1	AP1.2	AP2
1	Easy vs. standard wording of consent question	x	x	x	
2	Early vs. late placement of consent in questionnaire	x			
3	Web vs. face-to-face interview	x			
4	Default question wording		x		
5	Additional information question wording		x		
6	Data linkage domain				x
7	Trust priming				x
8	Multiple consents: format		x	x	
9	Multiple consents: order		x		x
	Sample sizes	2,896	5,684	2,064	3,850

Notes: IP = Innovation Panel, AP = PopulusLive online access panel.

Before analysing the experimental data, we checked that the random allocations to treatment groups had functioned as intended. For each of the experiments, there are no differences between treatment groups in terms of gender, age, education, whether in work, household composition, housing tenure, and country of residence.

Experiment 1: Easy versus standard wording of consent question

This first consent question wording experiment was designed to manipulate the difficulty of the request. Half of respondents were allocated to the 'standard' question wording, which had been used previously in the main *Understanding Society* survey. The other half were allocated to an 'easy' version, where the text was rewritten to reduce reading difficulty and to provide all essential information about the linkage in the question text rather than an additional information leaflet. The revisions were based on findings from our prior qualitative in-depth interviews about wording that hampered respondents' understanding of the consent request (Beninger et al. 2017) and on criteria used for reading level statistics. The randomisation for this experiment occurred within the survey, such that in IP11 the allocation to question wording was crossed with the mode in which respondents completed the survey. The wording of both question versions is documented in the appendix.

Experiment 2: Early versus late placement of consent question in the questionnaire

This experiment was crossed with the easy/standard question wording experiment and implemented only for face-to-face respondents. Half of the respondents were asked for consent early in the interview, the other half were asked at the end.

Experiment 3: Web versus face-to-face interview

The mode of data collection experiment was introduced in wave 5 of the IP. At the time, a random two-thirds of households were invited to complete the annual interview online (Web-first). Non-respondents were then followed up by face-to-face interviewers. The remaining third of households were allocated to face-to-face interviewers (FTF-first). By wave 11, non-respondents in the FTF-first group were followed up with invitations to complete the survey online. Across waves, households remained in the randomised groups to which they were initially allocated. From wave 8 onwards, however, households with very low predicted probability of completing the survey online were moved to the FTF-first group. All low web probability households (whether initially allocated to web-first or to FTF-first) are dropped from the analyses of how and why consent differs between modes of data collection (Section 4.3).

Experiment 4: Default question wording

The default version used the 'easy' request for consent to link to tax data. However, we dropped the yes/no question at the end of the question text, which explicitly asked the respondent *"Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?"* Instead, we presented linkage as the default, unless the respondent explicitly opted out. The respondent was instructed to *"Press 'next' to continue"* and given the option to click *"I do not want HMRC records to be added to my answers to this survey"*.

Experiment 5: Additional information question wording

This experiment was also based on the 'easy' consent question. In addition to 'yes' or 'no' as responses to the request for permission, we offered a third response option: "*I need more information before making a decision."*

Experiment 6: Data linkage domain

Respondents were either asked for consent to link their survey responses to tax data held by HMRC, or to health data held by the NHS. Both questions were based on the 'easy' consent question wording.

Experiment 7: Trust priming

This experiment was crossed with the data linkage domain experiment. For both HMRC and NHS consent we added an introductory screen saying, *"The next question is about linking the information you provide in this survey, to data that HMRC / the NHS hold about you."* Half the samples in both domain groups saw an additional statement: "*HMRC / The NHS is a trusted data holder"*. This was followed by an icon symbolizing data security: a shield and lock symbol with the heading *"Trust*" (see the question wording in the Appendix).

Experiment 8: Multiple consents: question format

This experiment manipulated how a series of five data linkage consent questions was presented to the respondent. For one group, the questions were each presented on a separate page. The respondent had to enter an answer and click 'next' before seeing the next question. For the second group the questions were all presented on one page, however the respondent still had to answer each consent question individually. For the third group all data linkage requests were presented on a single page and the respondent answered a single yes/no question, whether they consented to all the linkages or not.

Experiment 9: Multiple consents: question order

In the qualitative interviews (Beninger et al. 2017) respondents had expressed much concern about linking financial data held by the government to their survey data and expressed more acceptance of linkage to health data. In the first iteration of this experiment (in survey AP1.1) we had two question orders: one group was asked about HMRC tax data linkage, followed by DWP benefits and pensions data, BEIS household energy usage data, DfE education data, and NHS health data. The second group was asked the consent questions in reverse order, starting with the NHS data and ending with HMRC. In the second iteration of this experiment (in survey AP2) we replicated these two orders and added another two, by switching the order of the second and fourth consent question in both groups.

3.4 Behaviour coding of audio-recordings from face-to-face interviews

To examine what interviewers and respondents do when administering consent questions, we used digital audio-recordings of the IP11 face-to-face interviews. At the start of the interviews, respondents were asked for permission "*to record some parts of this interview to help us improve the questions we ask in future surveys"*. Overall, 86% of respondents gave permission and there were no significant differences in permission rates between the experimental consent groups. Based on the audio-recordings, we coded whether or not the following interviewer behaviours occurred:

1) whether the interviewer read the question as scripted or omitted parts of the question,
2) whether they handed over the information leaflet explaining data linkage,
3) whether they handed over or explained the diagram visualising the data linkage process,
4) whether they provided any additional information about the linkage,
5) whether they emphasised confidentiality, and
6) whether the interviewer gave an adequate response to concerns and questions raised by the respondent.

For the respondent, we coded:

7) whether or not they interrupted the interviewer and therefore did not hear the entire question text, and
8) whether they expressed uncertainty, concern, or asked a question.

4. Key findings

4.1 What effects do different ways of wording consent questions have?

In this section we first document the consent outcomes that we examine throughout, and how these relate to each other: whether the respondent consented, how well they felt they understood the data linkage request (subjective understanding), answers to knowledge test questions about the data linkage (objective understanding), and how confident the respondent was in their decision. These initial analyses are based on pooled data from respondents in the single consent treatment groups in the IP, AP1.1 and AP2 surveys, since the patterns are similar in each of the three surveys.

We then document the effects of the experimental manipulations of the single consent questions that are described in Section 3.3. These were designed to test specific hypotheses and are used in different ways in the analyses in the following sections. In this section, we document the effects of these manipulations on consent rates and on how well the respondent understood the linkage request, as measured by the test questions. In the graphs, "IPf" refers to face-to-face respondents from the Innovation Panel.

Key finding 1: Respondents who have better understanding of the data linkage request are also more likely to consent.

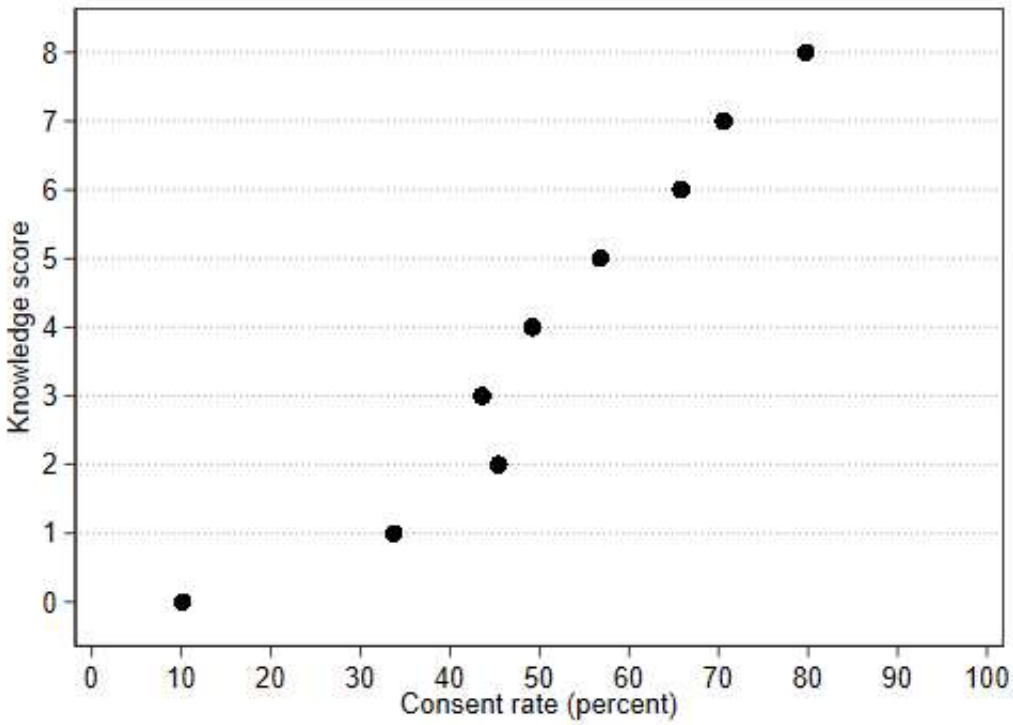
Figure 4.1.1 shows a strong and positive relationship between objective understanding of the request and consent. Among those who answered none of the knowledge test questions correctly, only 12% consented to the linkage; among those who answered half correctly 48% consented; and among those who answered all eight test questions correctly 72% gave consent. Key findings 3 and 6, however, suggest that this relationship is not causal. We return to this issue below.

The respondent's perception of how well they understood the request (subjective understanding) is also associated with consent, but not as strongly.

Subjective understanding is somewhat associated with objective understanding: the mean number of correct answers to the test questions ranges between 3.9 for those who said they did not understand the request at all, and 5.0 for those who said they completely understood.

Confidence in the consent decision is not at all associated with objective understanding of the consent request – and negatively associated with consent. While 65% to 71% of those who said they were not confident or somewhat confident in their decision gave consent, only 28% of those who were very confident consented.

Figure 4.1.1: Consent rate by number of correct answers to knowledge test questions



Key finding 2: As in previous studies, we find no socio-demographic characteristics that consistently predict consent in all samples. The only consistent predictors are positive attitudes towards data sharing, trust in HMRC, and knowledge of what data HMRC have.

In the background section of the questionnaire, respondents were shown three statements about sharing personal data and asked to what extent they agreed or disagreed with them. The statements were: "*I do not mind sharing personal information as nowadays everyone is doing this anyway"*, *"You cannot live in the modern world without sharing personal information"*, and *"I do not mind sharing personal information in return for a product or service that I want"*. The results show a statistically significant linear relationship between these attitudes and consent, that remains when controlling for sociodemographic characteristics of the respondent (see Appendix Table 1). To illustrate the effect size: at the extremes, respondents who 'strongly agreed' with each statement had a consent rate of 68%, compared to 28% among those who 'strongly disagreed'.

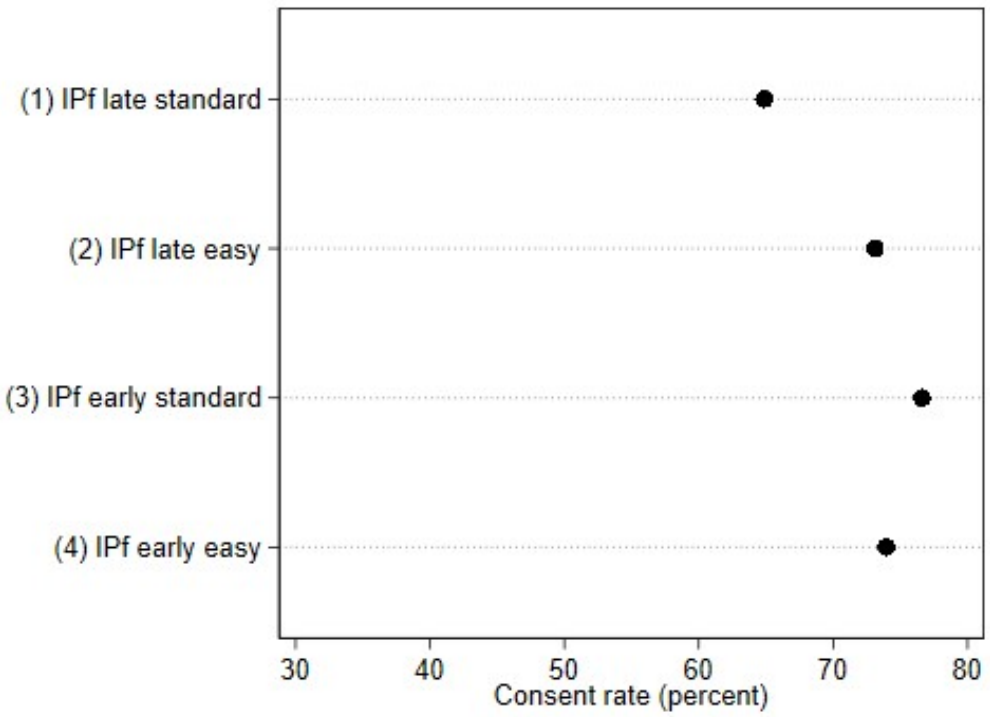
Respondents who said they trusted HMRC were more likely to consent than those who said they trusted HMRC only a little or not at all (58% versus 35%). This association also remains statistically significant after controlling for sociodemographic characteristics (see Appendix Table 1).

Respondents who said they knew, at least roughly, what data HMRC have about them had a 55% consent rate. In contrast those who did not know, or believed HMRC had no data about them, had a 44% consent rate. This association between knowledge of HMRC data and consent also remains significant when controlling for sociodemographic characteristics of the respondent (see Appendix Table 1).

Key finding 3: Respondents are less likely to consent to data linkage if the wording of the request is difficult and the question is asked late in the questionnaire. Position has no effect on consent if the wording is easy; wording has no effect on consent if the position is early.

Rows 2 to 4 in Figure 4.1.2 show that the consent rate ranged between 73% and 77% if the standard wording was asked early, or the easy wording was asked (early or late). In contrast, if the standard question was asked late (row 1), the consent rate is 65%. That is, there is a statistically significant interaction between wording difficulty and late position of the consent request (see Appendix Table 1).

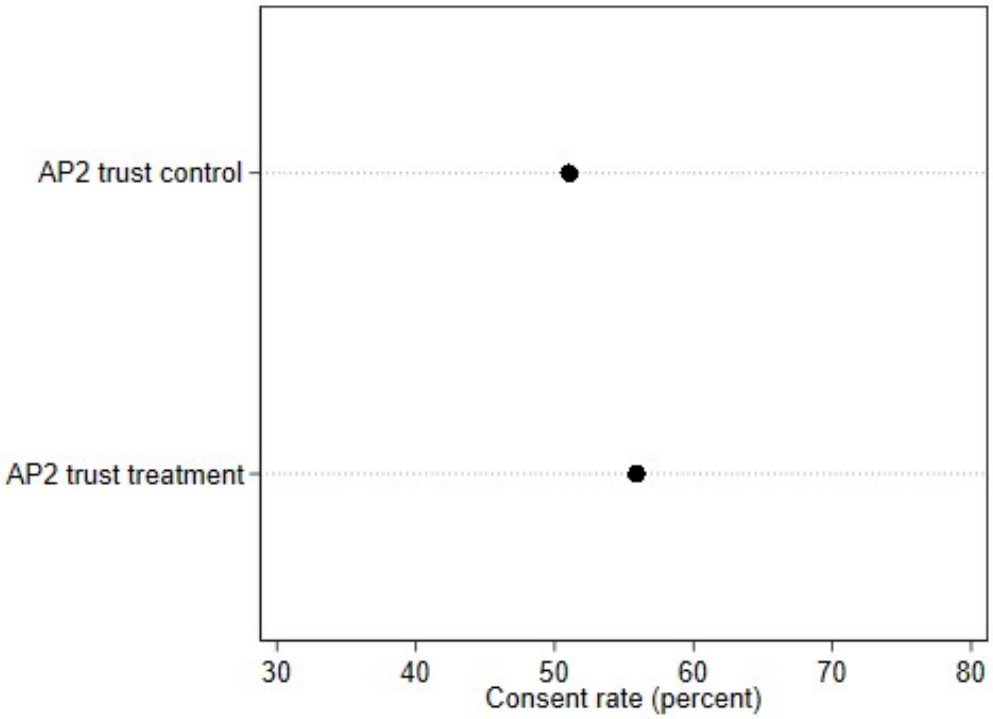
Figure 4.1.2: Consent rate by question wording and position



Key finding 4: Priming respondents to think about trust in the organisations involved in the data linkage increases consent.

Figure 4.1.3 shows that respondents who were primed to think about trust in the organisations involved had a consent rate of 56%. The control group had a consent rate of 51%. This treatment effect is statistically significant (see Appendix Table 1).

Figure 4.1.3: Consent rate by trust treatment group



Key finding 5: The only socio-demographic characteristic that consistently predicts objective understanding of the linkage request is education*. *Understanding is positively associated with the number of online data sharing behaviours and with trust in HMRC.

Respondents in the lowest education group scored on average slightly lower (about half a point on a scale with a possible score ranging from 0 to 8) on the test questions than those with a degree. This effect remains statistically significant after controlling for other characteristics (see Appendix Table 2).

Trust in the HMRC was also positively associated with understanding: those who trusted HMRC scored between .3 and .4 points higher than those who did not trust HMRC (see Appendix Table 2).

Respondents who share personal information online had higher understanding of the request than those who do not. Respondents were asked: *"Which of the following do you do? 1) post text on*

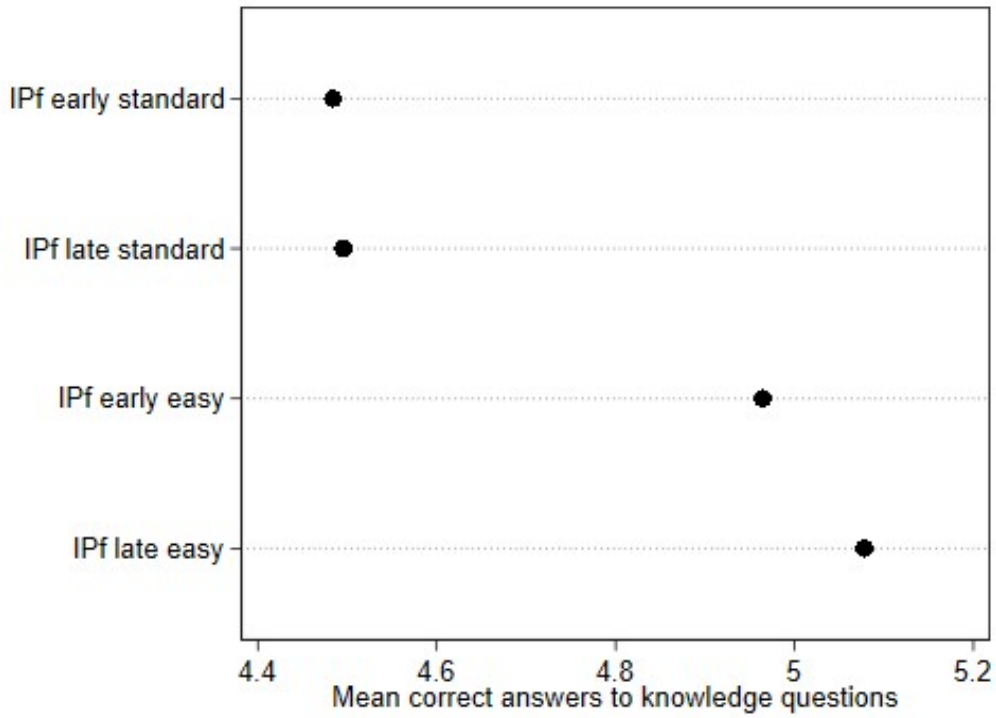
social media websites or apps, 2) post images of yourself, family or friends online, 3) post videos online, 4) download apps onto your smartphone or tablet, 5) make purchases online (for example booking train tickets, buying clothes, ordering food), 6) use online banking (for example checking account balance, transferring money), 7) use GPS/location-aware apps (for example Google Maps), or 8) none of these". Respondents who do none of these data sharing behaviours answered on average 3.9 of the test questions correctly. Those who do four or more of the behaviours answered on average 4.7 questions correctly (see Appendix Table 2).

While the effect of education on understanding of the data linkage request is plausibly a causal effect, the effects of trust in HMRC and of data sharing behaviours are unlikely to be causal effects.

Key finding 6: Easy wording of the consent question increases objective understanding of the linkage request. Position has no effect.

Figure 4.1.4 shows that respondents who received the easy question wording answered on average around 5 of the test questions correctly. In contrast, those who received the standard wording answered around 4.5 questions correctly. This effect of question wording on objective understanding is statistically significant (see Appendix Table 2).

Figure 4.1.4: Objective understanding of the consent request by question wording and position



Implications for practice

The findings presented in this section highlight the importance of trust: respondents with higher levels of trust in the organisations involved are more likely to consent. Experimentally priming respondents to think about trust in the organisations increases consent.

Easier wording of the consent request – better readability and providing all relevant information in the question rather than in additional information materials – increases understanding of the request and can increase consent.

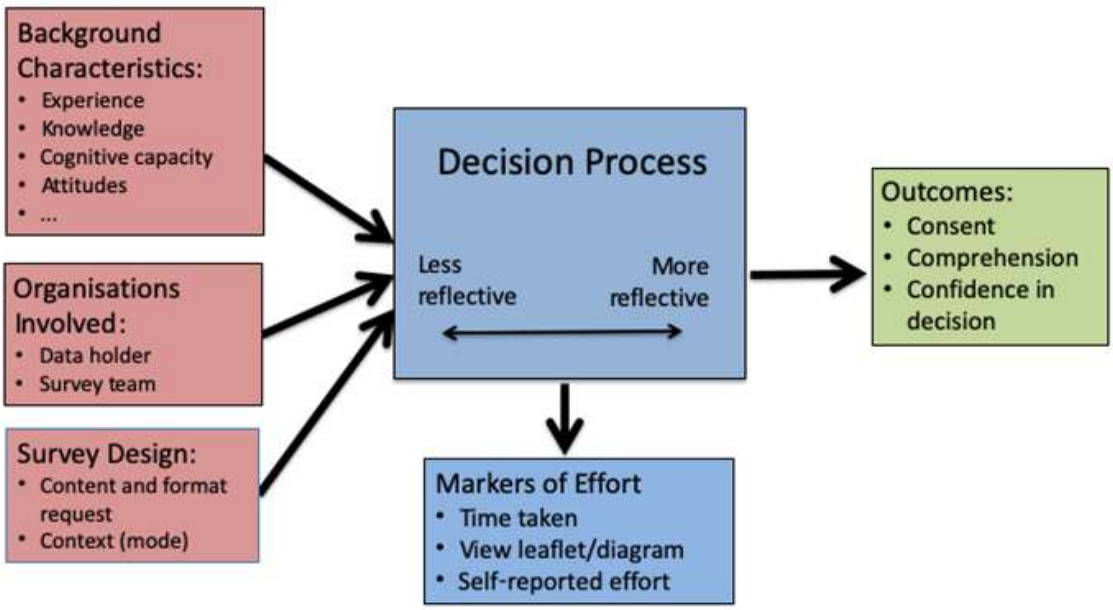
Future directions

The findings summarized in this section suggest a puzzle: while understanding of the linkage request is strongly associated with consent, this does not appear to be a causal link. When we presented respondents with an easier version of the consent question, this improved their understanding but did not increase consent. Similarly, when we asked them later in the questionnaire, this reduced the consent rate but did not affect understanding. The question then is, what drives the association between understanding and consent?

4.2 How do respondents make the decision whether or not to consent to data linkage?

The conceptual framework that guides our analysis in this section is summarized in Figure 4.2.1. There are two key ideas. The first is that respondents differ in how they make the consent decision. All respondents use some kind of "fast and frugal" or heuristic decision processes, but those decision processes range from more reflective (drawing on incomplete information about the consequences of consent, or the survey organization or data holders) to less reflective (more instinctive, or habit-based).

Figure 4.2.1: Conceptual framework



The second idea is that the decision process mediates the effect of background variables and survey design elements on outcomes including the decision itself, and comprehension of the request. At the same time, we do not rule out that characteristics of the respondent, survey organization, consent request and the survey context may have direct effects on these outcomes. An important implication of this is that it is very challenging to estimate the causal effect of the decision process on outcomes, a point that we return to below.

There are many factors that could affect the processes survey respondents use when making a decision about consent. Some of these are individual characteristics, experiences and attitudes that may predispose respondents to a certain decision process or even outcome. These may include such factors as cognitive capacity, prior knowledge or experience of administrative data and risks of linkage, motivation, general attitudes towards data sharing and privacy, and so on. A second set of

factors that may influence the decision process adopted relate to attributes of the organizations involved, including both the survey organization and the data holder. Participants' knowledge of, experience with, and trust in, those organizations may influence how they process the request. Finally, the design of the consent request and the survey context in which it is posed, may affect the decision process. Both the content (what information is conveyed) and the format (how the information is presented) of the request are important here. The survey mode in which the request is delivered and the role of the agent (interviewer) in delivering the request also play a role. Most of the research focus on consent has manipulated the content and format of the request (e.g., gain versus loss framing, longer versus shorter descriptions of the linkage process, etc.).

Key finding 7: Respondents decide about the consent request in different ways.

Our first finding is that respondents give heterogeneous responses when asked, after the fact, *how* they made the consent decision. This is illustrated in Figures 4.2.2 (Innovation Panel samples) and 4.2.3 (Access Panel samples). Those who chose the response *I thought about what would happen if I said "yes" or "no"* we label "consequentialist," and they are just 30-40% in each sample. Other respondents used decisions processes that were less reflective (more instinctive). For example, across all samples, about 30% of respondents selected the option "instinct or gut feeling" (labelled "gut feeling" in the Figures). Note that multiple response options were allowed, but these were infrequent in the Innovation Panel samples.

In the AP samples, we added the response option *I thought about how much I trust the organisations involved* (labelled "trust" in Figure 4.2.3). This is the most frequently chosen response option in these samples. Note also that in these samples it was more common for respondents to choose more than one response option. The most common combination was "consequentialist" and "trust".

In sum, respondents tell us that they made the decision in different ways. A significant fraction of respondents report thinking about the consequences of consent, but another group instead consider their trust in the survey organization and data holder, and further groups make less reflective decisions, relying on a gut feeling or instinct, or on habit or usual practice.

Figure 4.2.2: Self-reported decision process, IP samples

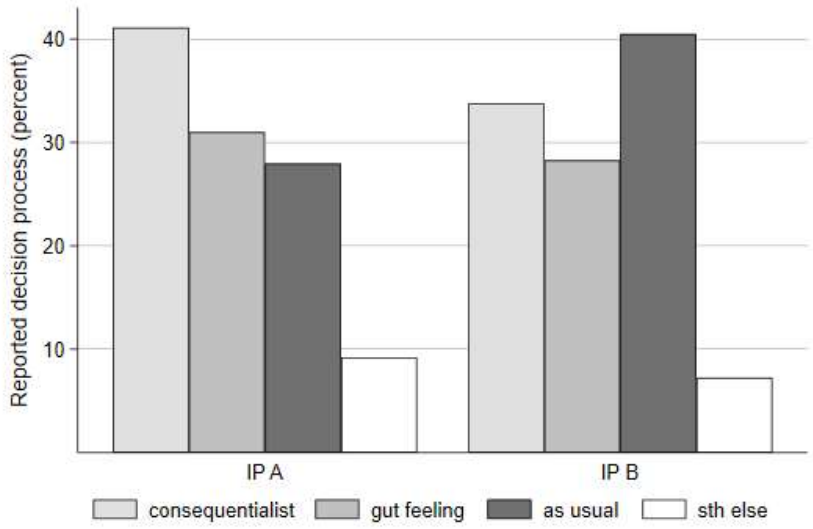
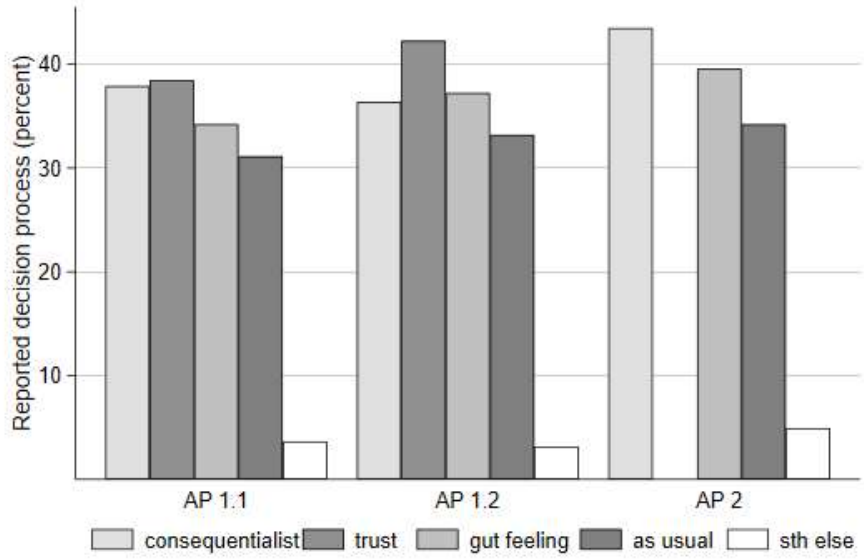


Figure 4.2.3: Self-reported decision process, AP samples



Are these self-reports of decision process meaningful? To assess this question, we looked at markers of decision effort and these are quite consistent with the self-reports. For example, in the AP1.1 sample, respondents who reported “consequentialist”- or “trust”-based decision processes had median response times for the consent question of 48 and 59 seconds respectively. In contrast, those respondents who reported “gut feeling” or “as usual” decision processes had much shorter response times, with medians of 24 and 28 seconds respectively. Qualitatively similar results were obtained in all samples. In addition, in the AP1.2 sample we asked a self-reported “decision effort” question (on an 11-point scale) and again the responses align with the chosen descriptions of the

decision process, with those reporting “consequentialist” or “trust” reporting greater effort than those reporting “gut feeling” or “as usual”. Finally, in the AP2 sample we also asked people how much of a role different considerations played in making their decision. Again, responses to this question are consistent with the labels that respondents chose for their decision process. This is shown in Table 4.2.4. Each item was rated on a scale five-point scale where “1” was labelled “played no role” and “5” was labelled “Played a very big role”. Items such as “what information the government has about me”, “how much I know about the organizations involved”, and “the benefits to society” played a larger role for those respondents who reported a consequentialist decision process.

These findings give us confidence that the self-reported decision processes measure genuine differences in how the respondents arrived at their answer to the consent question.

Table 4.2.4: What did respondents think about when making consent decision?

	consequentialist	gut feeling	as usual
Information			
What information the government has about me	3.6	3.1	3.1
How much information I've already shared	3.3	2.9	3.2
How much I know about the organisations involved	3.7	3.5	3.2
Risks			
Who might have access to my information	4.1	3.7	4.0
How much I trust the organisations involved	4.0	3.7	3.5
Whether the Government would use this data to check on me	3.5	3.2	3.0
Recent news stories related to data security	2.6	2.8	2.9
The chances of my personal data getting into the wrong hands or being misused	3.9	3.6	4.0
Benefits			
Wanting to be helpful	3.2	3.0	2.6
The benefits to society	2.9	2.5	2.3
Processes			
The way I usually make decisions	3.3	3.4	3.8
My gut feelings	3.1	4.2	3.7
What friends and family would do	1.9	2.2	2.1
Avoiding difficult decisions	2.3	2.7	2.3

Notes: mean scores from five-point response scale labelled 1 = played no role, 5 = played a very big role.

Key finding 8: Different decision processes are associated with very different levels of consent, comprehension, and confidence in the consent decision.

The core of our analysis is to examine the associations between self-reported decision processes and key outcomes: the consent decision itself, and comprehension of the request. Table 4.2.5 shows that respondents who reported "consequentialist" or "trust"-based decision processes had higher consent rates than those who reported less reflective processes ("gut feeling" or "as usual"). The differences are large, with the consent rate in the "consequentialist" or "trust" groups often double or more the consent rates in the "gut feeling" or "as usual" groups. The differences are also statistically significant at conventional levels. In those samples where we offered "trust" as a response option, it is associated with the highest consent rates. Those who report "as usual" have very low consent rates.

Table 4.2.5: Consent rate by reported decision process

	IP11f	IP11w	AP1.1	AP1.2	AP2
Consequentialist (%)	85.2	66.5	54.6	52.9	69.5
Trust (%)	-	-	83.2	88.7	-
Consequentialist + trust (%)	-	-	77.1	82.8	-
Gut feeling (%)	75.0	48.5	42.9	40.8	53.4
As usual (%)	52.7	19.6	19.5	19.8	24.8

Comprehension of the consent question (Table 4.2.6) follows a similar pattern. Recall that the measure of comprehension is the sum of correct answers to 8 knowledge questions about the information provided about the data linkage for which consent was sought. Respondents who reported "consequentialist"- or "trust"-based decision processes had better comprehension than those who reported less reflective processes ("gut feeling" or "as usual") The magnitude of the differences are 0.5-1.0 correct items (in means), and the differences are again statistically significant at conventional levels.

Table 4.2.6: Comprehension by reported decision process (mean score of objective measure)

	IP11f	IP11w	AP1.1	AP1.2	AP2
Consequentialist	5.2	4.9	4.5	4.6	5.0
Trust	-	-	4.8	4.7	-
Consequentialist + trust	-	-	5.2	5.1	-
Gut feeling	4.7	4.2	4.1	4.2	4.4
As usual	4.2	3.8	4.1	4.1	4.2

Key finding 9: Placing the consent request earlier in the survey increases the probability of the respondent using a reflective decision-making process.

Finally, we examined determinants of the decision process reported. Among background characteristics of the respondent, education seems important. Those with a degree were more likely to report a "consequentialist" decision process, for example (and this was statistically significant at conventional levels in all but one sample.) The effects of the placement and drafting of the consent request are of particular interest as they are controlled by the survey designer. Recall that in the IP11f sample we experimentally manipulated both the placement in the survey of the consent question (early or late) and readability of the question. We find that placing the question early in the survey increases the probability that the respondent uses a "consequentialist" decision process by 10 percentage points (against an overall rate of about 40%), and this effect is statistically significant at conventional levels. We also find an effect of the "easy wording" of similar magnitude, but only with the late placement of the question.

Implications for practice

These findings suggest that survey respondents formulate a response to a data linkage consent question in a variety of ways. Some take a more thoughtful decision processes, reflecting on the possible consequences of giving consent, or on their degree of trust in the organizations involved in the request. Others decide in more rapid, less reflective, ways, going with a "gut instinct" or a "usual" response. It is important to note that even the more reflective decision makers are not making "fully rational" decisions. They necessarily have incomplete information about the consequences of consent, and even the slowest decision process produced responses in, on average, less than a minute.

Nevertheless, different decision processes are associated with very different levels of consent and comprehension, with the more reflective decision processes associated with better outcomes on both measures. More reflected decision processes are associated with more consent and more informed consent.

Many attempts to achieve higher consent rates and more informed consent involve the provision of additional information, and (as reviewed above) these have had limited success. Our results point to an explanation: if many of those who withhold consent are using a very rapid and unreflective decision process, additional information is unlikely to be incorporated into their decision, and hence unlikely to change that decision.

Instead, our results suggest that a fruitful strategy for promoting informed consent may be to try to shift respondents towards more reflective decision processes, whether that be a "consequential" decision process, or one based on trust in the relevant organizations.

Future directions

It would obviously be desirable to establish that the associations between decision processes and consent outcomes we have document are in fact causal effects. To do so requires a source of variation (e.g., an instrumental variable) that affects outcomes *only* through the decision process. It is not immediately clear what factors would satisfy this requirement.

We can of course test for the effects of experimentally manipulated survey design features on consent outcomes. Our results suggest new initiatives that could be tested. These include further design manipulations to encourage more reflective decision processes.

A further direction to investigate is whether we are providing the right kind of information. For instance, much of the focus of information provision associated with consent questions is on the *process* of linkage, which may be irrelevant to respondent concerns about access to and use of the linked data. While it is much harder to describe the risks of linkages for respondents (see related work on survey disclosure risk by Couper et al. 2008, 2010; Singer and Couper 2010), this may be of more direct relevance to the respondent's decision process. Alternatively, given the prevalence of trust-based decision processes, information about the organizations involved may be more relevant than the process of linkage per se.

4.3 How does the mode of interview influence consent to data linkage?

As outlined in Section 2.3, one consistent finding in the literature is that consent to administrative record linkage is lower in web surveys than in face-to-face interviews. Given the rising use of the web for data collection and increased demand for administrative record linkages, there is a pressing need to understand the reasons behind these observed mode differences, and to find ways to increase rates of consent, especially on the web.

In this section we report results from the mode experiment in the Innovation Panel. Using Figure 4.2.1 as a guiding framework, we examine differences between modes in consent outcomes, in how respondents process the consent request, and in background factors that determine consent.

Our hypothesis is that differences between modes of data collection – such as whether an interviewer is involved and present, the speed at which respondents answer the survey questions, and the technology used to administer the questionnaire – can alter factors that influence the respondent's consent decision.

Key finding 10: As in previous studies, respondents are less likely to consent online than with an interviewer.

Our study replicates the results of previous studies, showing that consent rates online were lower than in-person. Although respondents in the IP were randomly allocated to mode treatment groups (FTF-first or web-first), the actual mode of interview is not randomised due to non-compliance of respondents with the initial mode assignment. Therefore, the differences we observe in the web group and the face-to-face group may be due to either measurement effects (the mode affects how people respond) or selection effects (different people take part in each mode). We therefore examine the effects of the data collection mode in three ways: using the mode of interview (as-treated analysis), using the randomised allocation to mode (intention-to-treat analysis), and using the randomised allocation to mode as an instrument for the actual mode of interview (instrumental variable analysis).

Table 4.3.1 below shows how consent, subjective and objective understanding of the linkage request, and confidence in the consent decision differed between modes. Using the mode of interview, we observe that web respondents consented at a lower rate (42.0%) than face-to-face respondents (72.8%), a difference of 30.8 percentage points. The intention-to-treat analysis suggests that only part of this difference between modes was due to different types of respondents selecting into web and face-to-face interviews: the difference between mode allocation groups remained large at 21.6 percentage points. Scaling the intention-to-treat estimate up to estimate the effect of

mode if all respondents completed the survey online (using the instrumental variable analysis), suggests that the consent rate would be 30.4 percentage points lower than in if all respondents completed the survey with an interviewer.

Table 4.3.1: Consent outcomes by survey mode (Innovation Panel)

	Mode of interview		Mode allocation	
	FTF	Web	FTF-first	Web-first
Consent (%)	72.8	42.0	70.6	49.0
Completely/mostly understand (%)	65.7	43.5	64.1	48.5
Mean number of correct answers	4.8	4.1	4.8	4.2
(Very) confident in consent decision (%)	73.0	71.6	73.8	71.3

Key finding 11: Web respondents have lower levels of understanding than face-to-face respondents.

Examining how well respondents felt they understood the linkage request, 65.7% of face-to-face respondents said they completely or mostly understood the consent request, compared to 43.5% of web respondents (a difference of 22.2 percentage points). The intention-to-treat analysis again suggests that the difference was only partly due to selection into mode: the difference remained at 15.6 percentage points. This question was part of the self-completion section for those interviewed in-person, and so we do not think that the higher rates of self-reported subjective understanding can be explained by social desirability bias.

The objective measure of understanding based on the knowledge test questions paints a similar picture: web respondents were more likely to be at the low end of the understanding score, with 18.8% answering 0-2 of the true/false knowledge test items correctly, compared to 8.1% of face-to-face respondents. Further analyses again suggest that the difference is due to the effect of mode on responses and not due to selection into mode.

Key finding 12: There is no difference by mode in respondents' confidence in their decisions.

Although consent rates and understanding of the consent request were lower in web than face-to-face interviews, there is no difference between modes in how confident respondents were about their consent decision. Across modes, nearly three-quarters of respondents said they were confident or very confident in their decision.

Key finding 13: Web respondents report higher levels of concern about data security than face-to-face respondents.

Web respondents reported higher levels of concern about privacy and data security than face-to-face respondents. Among face-to-face respondents 62.8% said they were very or somewhat worried about privacy and 78.0% said they were very or somewhat concerned about data security. The corresponding percentages among web respondents were 71.1% and 83.1%. These differences again remain significant when accounting for selection into modes.

There was no difference between modes in respondents' perception of how sensitive HMRC data are or their trust in the organisations involved in the data linkage.

Key finding 14: Web respondents are less likely to use reflective strategies to make their decision than face-to-face respondents.

Table 4.3.2 examines differences between modes in how respondents process the consent decision. Web respondents were less likely than face-to-face respondents to think about what would happen if they said yes to the consent request (26.5% compared to 35.3%). Web respondents were instead more likely to make a habit-based decision about consent (34.1% compared to 22.0%). These differences in decision processes between modes remain significant when accounting for selection into modes.

Table 4.3.2: Consent decision process by survey mode (Innovation Panel)

	Mode of interview		Mode allocation	
	FTF	Web	FTF-first	Web-first
Decision process:				
Consequentialist (%)	35.3	26.5	34.3	28.7
Gut feeling (%)	25.6	22.1	25.2	23.0
As usual (%)	22.0	34.1	23.4	30.9
Something else or combinations (%)	17.1	17.3	17.1	17.3
Time (mean seconds)	94.7	42.9	91.2	54.5
Read/clicked on leaflet (%)	40.1	8.4	37.9	15.5
Amount of information too much (%)	8.4	15.7	9.2	13.9

These differences between modes are also reflected in other measures of how respondents processed the consent request. Web respondents answered the consent question more quickly than face-to-face respondents: the average response times were 42.9 seconds versus 94.7 seconds. Some of the difference in response times is also due to the fact that face-to-face respondents were more likely than web respondents to view the additional materials explaining data linkage. The interviewer

observations indicate that 40.1% of respondents read the information leaflet. In contrast, only 8.4% of web respondents clicked on the link to the leaflet. Finally, web respondents were more likely to say that the amount of information presented about data linkage was too much (15.7% compared to 8.4% of face-to-face respondents). These differences between modes also remain after accounting for differences in selection.

Key finding 15: Easier wording of the consent request does not reduce mode effects.

We had hypothesized that easier wording of the consent request might help online respondents with processing the consent request. However, while the easier wording increased objective understanding, it had the same effect in both modes of interview, and no effect on consent.

Key finding 16: Respondents rarely ask questions and interviewers rarely provide additional information.

Analyses of the audio-recordings of face-to-face interviews show that few respondents make use of the opportunity to ask the interviewer questions: only 16.3% asked a question and 5.3% expressed concern or uncertainty. Respondents were equally likely to ask questions or express concern if the wording of the consent question was easy rather than standard, and if it was asked early rather than late in the questionnaire. That is, even though respondents understand the consent request less well, when asked the standard wording rather than the easy wording, they do not make use of the opportunity to ask the interviewer for clarifications.

Similarly, interviewers rarely volunteer additional information to explain the data linkage request or offer reassurance: they emphasized confidentiality in only 4.1% of cases. They were however significantly more likely to provide additional unsolicited information with the standard consent question wording (17.3% of cases) than with the easy wording (10.9%).

Implications for practice

These findings indicate that there is a real effect of the mode on how respondents make the consent decision, which is not explained by selection into mode. Web respondents are less likely to use reflective strategies in their decision-making process, and so respond more quickly and, although they have confidence in their decision, are less likely to have understood the consent request. This implies that providing additional information in online surveys is unlikely to increase informed consent, because respondents are not likely to read it.

Future directions

The results suggest that higher consent rates in face-to-face interviews are unlikely to be due to interviewers providing additional information, clarifications and reassurance. The results are however consistent with other potential explanations for why consent rates are higher in interviewer administered surveys. First, the presence of an interviewer means that how the interview unfolds is influenced by social norms, which might mean that respondents are reluctant to decline the request. Second, the faster speed with which online respondents complete questionnaires might be a reason why they process the consent request less reflectively. Third, the technology used to administer the questionnaire (an interviewer's laptop versus an internet browser) might prime respondents to be more or less concerned about the security of their data.

4.4 How best to ask for consent to multiple data linkages?

As noted in Section 2.4, most of the research on data linkage consent has focused on single consent questions. There is evidence of variation among domains within a single survey, and some evidence of order effects. Given this and the rising demand for linkages to multiple administrative data sources, it is important to understand how best to ask consent to several domains in a single survey. There are many ways in which this could be done. For example, the requests could be spread throughout the survey, coupled with substantive questions on the relevant domain; or they could be asked together in a set. Given that most surveys ask consent at the end of the survey, our initial research focuses on the latter approach. We focus on addressing two research questions when presenting multiple requests for consent: 1) does the order matter, and 2) does the presentation format matter.

With regard to order, the literature on sequential requests in social psychology (see Dillard 1991) suggests that the order in which requests are made affect the compliance with later requests. A common mechanism identified in this literature is the foot-in-the-door (FITD) effect. FITD can be traced back to the research on "compliance without pressure" by Freedman and Fraser (1966, p. 195) who posited that "once a person has been induced to comply with a small request, he is more likely to comply with a larger demand." In contrast, the door-in-the-face (DITF) effect posits that "a request made is more likely to be agreed to if preceded by the offer and refusal of a more expensive one" (Cialdini et al. 1975). See Dolinski (2011) for a fuller discussion of these two mechanisms. However, this literature focuses on maximizing consent to a single target request and is not directly relevant to our research. We thus expect order to matter, but don't have strong priors over whether it is better to start with the least sensitive or risky request, or to start with the most sensitive request. Our goal is to investigate what carryover effects occur when asking about multiple consent domains.

Turning to the issue of presentation format, there is research in other areas of survey research that suggests that how questions are grouped together affect the context in which they are processed (see, e.g., Schwarz and Sudman, 1992; Tourangeau, Rips, and Rasinski, 2000). Questions grouped together on the same page are seen as a single element and processed jointly. Questions on separate pages are processed sequentially. Similarly, asking respondents to consent to each of a set of different domains in turn may impose more demands on them than asking a single consent question about all five domains together.

We have several outcomes of interest. The strategy used may depend on whether the goal is

1) to maximise the average consent rate across the domains, or
2) to minimise the share of respondents who do not agree to any type of linkage, or
3) to maximise the consent rate in a specific consent domain.

We thus examine all three of these outcomes.

Study 1

In our first study we tested consent to five different domains, varying both the format of the request (Experiment 8 in Table 3.3.1) and the order (Experiment 9 in Table 3.3.1). The five domains are: tax data (held by HM Revenue and Customs, HMRC), health data (held by the National Health Service, NHS), education data (held by the Department for Education in England, DfE, and equivalent departments in Scotland and Wales), household energy data (held the Department for Business, Energy and Industrial Strategy, BEIS), and benefit and pensions data (held by the Department for Work and Pensions, DWP). Analysis of prior consent rates to these five domains suggests that NHS is the most acceptable request (produces the highest consent rates) and HMRC the least (lowest consent rates).

We conducted the experiment in the PopulusLive online access panel (AP1.1). Given the complexities of the experimental design and the limited available sample in the *Understanding Society* Innovation Panel (IP11), we made this decision knowing that while external validity (i.e., the *rates* of consent to the various domains) may be limited, we should still have strong internal validity (the *differences* in rates between the different experimental conditions). A full randomization of all sequences of five consent domains would have produced too many conditions, so we limited our experiment to two sequences. This design yields a 2 (order) by 3 (format) experiment as follows:

Order:

- HMRC – DWP – BEIS – EDUC – NHS (“HMRC-first”)
- NHS – EDUC – BEIS – DWP – HMRC (“NHS-first”)

Format:

- Sequence of pages (with one response per domain)
- Same page (with one response per domain)
- Single request (with one joint request covering all 5 domains)

We have over 500 respondents in each of the six experimental cells.

Key finding 17: The format in which a sequence of consent requests is asked does not seem to matter.

To address this question, we look at both average consent rates and the percentage of respondents who answer yes or no to all five requests. Figure 4.4.1 shows the average consent rate by both format and order. While it is clear that order matters (lower average consent rates when HMRC is asked first), the format of presentation has no effect on average consent rates at conventional levels of statistical significance.

Figure 4.4.2 shows the percentages of respondents saying yes to all 5 requests, no to all 5 requests, or a mix of responses. By definition in the single question format there are no mixed responses.

Still focusing on the format manipulation and looking at the "yes to all" group, there is no difference between the sequence of pages (first pair of bars) and same page (second pair). The single request has a higher rate of "yes to all" than the other two groups. However, it also has a higher rate of "no to all". Respondents who are unable to give a mixed answer in this group appear to distribute their answers roughly equally between these two responses. One the one hand, this suggests that asking a single yes/no consent question does not produce higher rates of non-consent. This is potentially good news, considering that respondents in the qualitative study (Beninger et al. 2017) had expressed a preference for separate questions for each domain over such a "catch all" or joint question. On the other hand, a sizeable minority of respondents (ranging from 12.9% to 23.7%) differentiated between domains in their consent decisions, pointing to the potential need to provide respondents with this opportunity. Further, individual data holders may not accept such a joint decision request for ethical or legal reasons. Our conclusion is thus that the format of presentation of multiple consent domains has little impact on the consent rates.

Figure 4.4.1: Average consent rates by experimental condition

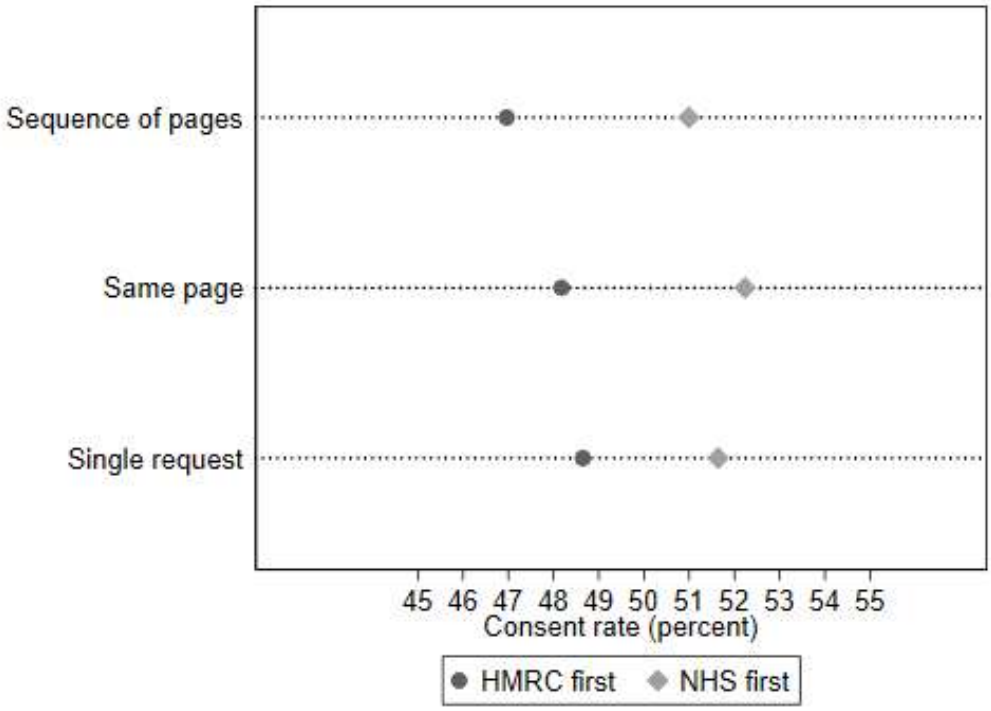
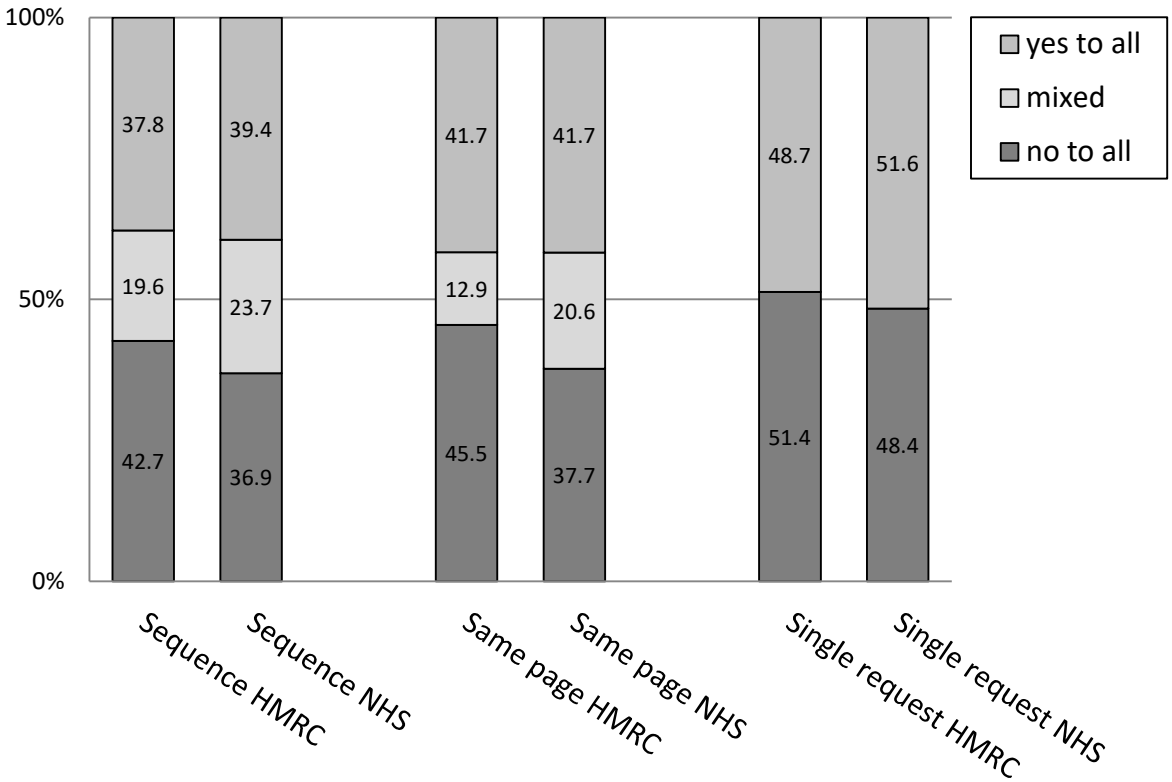


Figure 4.4.2: Yes-to-all / no-to-all requests by experimental condition

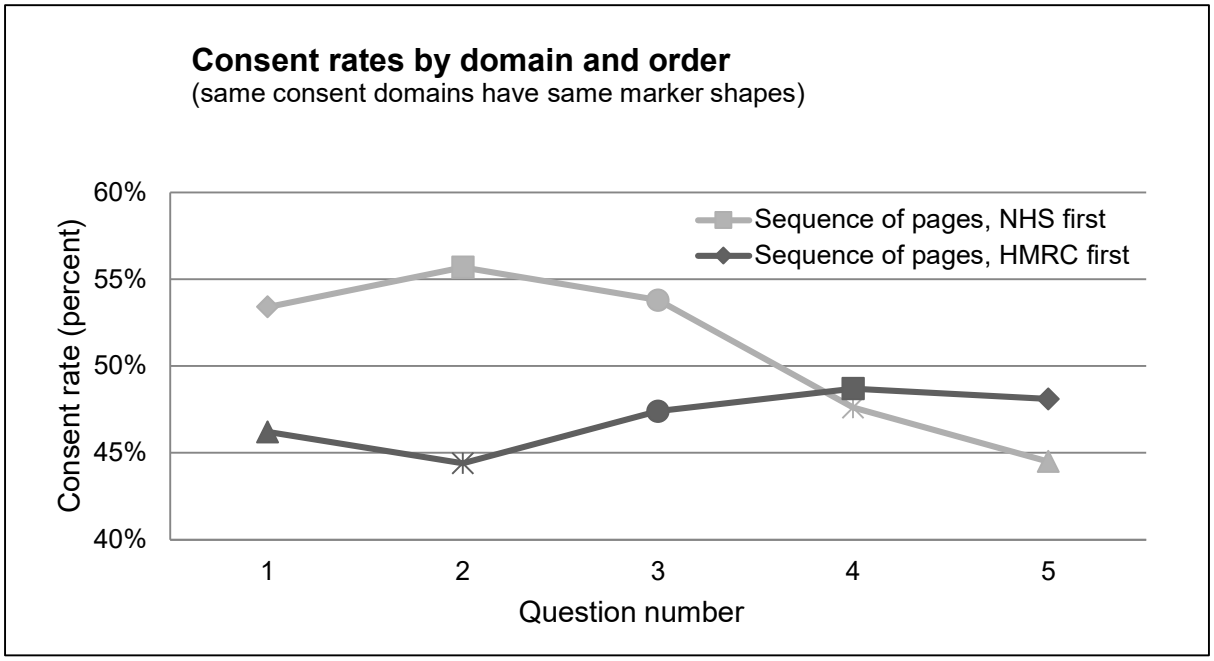


Key finding 18: The order of multiple consent requests affects consent rates.

It is also clear from the two figures above that the order in which the domains are asked affects consent rates. But this depends in part on what outcome is being examined. If we look at Figure 4.4.1, it is clear that asking the more acceptable consent request (NHS) first increases the average consent rates across the five domains by 3.6 percentage points. Similarly, Figure 4.4.2 shows that asking consent to NHS first increases the percentage of yes to all slightly (1.6 percentage points) in the sequential condition, but not in the same page condition. On the other hand, both conditions reduce the percentage of "no to all" when NHS is asked first. Consistent with the foot-in-the-door hypothesis, this suggests a small benefit of starting with the most acceptable request when making a sequence of consent requests.

Another way to look at this is to focus on the effect of order on individual consent rates. To make this easier to see, Figure 4.4.3 includes only the sequence order condition. The grey line is for the NHS first order, while the black line is for HMRC first. Markers of the same shape are for the same domain.

Figure 4.4.3: Individual consent rates by domain and order, sequence of pages



We see a "scissor" pattern in the graph. Asking NHS consent first yields a higher rate of consent to NHS linkage than asking NHS last (difference: 8.9 percentage points). Asking HMRC consent first yields only a slightly higher rate over asking it last (difference: 1.9 percentage points). BEIS consent is in a constant position (3rd), so we can examine whether consent rates to BEIS are affected by order. We can see from the graph that asking NHS first yields a higher consent rate to BEIS: a 6.4

percentage point difference in the sequential condition (as seen in Figure 4.4.3) and a 3.9 percentage point difference in the same page condition (not shown).

Our conclusion is that the order in which multiple consent requests are asked matters, but in complicated ways that depend on the particular outcome in which one is interested.

Key finding 19: Objective knowledge, subjective understanding and subjective confidence in the decision do not differ much by order and format.

As noted throughout this report, a key goal is not just maximising the rate of consent, but also ensuring that such consent is informed, that is, that objective knowledge of the consent process, subjective understanding (how informed respondents feel) and confidence in the consent decision are not affected by the experimental manipulations. We find that order of the consent requests has no effect on objective knowledge scores and subjective understanding. The sequential format is associated with slightly higher knowledge scores than the single page and single request formats. On the other hand, subjective understanding is slightly higher in the single request format, while confidence does not differ by order or format. These relatively small and inconsistent effects suggest that order and format do not affect objective and subjective understanding and confidence in reliable ways; that is, maximising consent rates does not come at a cost of reduced knowledge or confidence in the decision.

One final factor to consider is how long it takes to make the consent decision. Across both orders, we find that the sequence of pages takes the longest (an average of 71.6 seconds), followed by the single page version (67.1 seconds) and then the joint request version (44.6 seconds). This suggests that in the joint request condition (with a single yes/no question), respondents are not processing the individual consent requests as thoroughly as when the requests are presented separately.

Study 2

Given the suggestive findings on the effects of question order in the first study, we conducted a follow-up study in which we expanded on the sequences of question orders. Remember than in study 1, only BEIS was asked in the same position (third) in both orders. We increased the number of conditions so that we could also examine the effects of order on consent to DWP and EDUC linkage (both now asked either 2nd or 4th). Thus, two of the conditions replicated the orders used in Study 1, but we added another two order conditions. This results in the following four orders:

- HMRC – DWP – BEIS – EDUC – NHS
- HMRC – EDUC – BEIS – DWP – NHS
- NHS – DWP – BEIS – EDUC – HMRC

- NHS – EDUC – BEIS – DWP – HMRC

Given that we found little consistent effect of format, we tested these four orders using the sequence of separate pages. The second study was again implemented in the PopulusLive online access panel (AP2).

Key finding 20: The order effects from Study 1 do not replicate in Study 2.

The first surprise is that the two conditions that were exact replications of Study 1 produced different results in Study 2. The average consent rates do not differ much across the first three groups (ranging from 49.1% to 50.1%), but the 4th group (NHS first) had a noticeably lower average consent rate (45.7%). This is in sharp contrast to Study 1, where this condition had a higher consent rate than the first condition (HMRC first). The fourth group also yielded a lower rate of "yes-to-all" (34.3%) than the other groups in Study 2 (40.2% for the third group, also with NHS first), but also lower than the corresponding group in Study 1 (39.4%).

The second surprise is that the results for BEIS consent also do not replicate. Recall that BEIS has a 6.4 percentage point higher consent rate when following NHS consent than when following HMRC consent. In Study 2 this difference trends in the opposite direction (1.2 percentage points lower when following NHS) but does not reach traditional levels of significance. That is, we no longer find a positive carryover effect on BEIS consent of asking NHS consent first. The results for the other two consent domains (DWP and EDUC) also do not show a consistent pattern by order.

The differences between the consent rates for NHS and HMRC when each is asked first (i.e., unaffected by subsequent domains) are in the expected direction (higher for NHS) but not as large as observed in the first study and fail to reach traditional levels of significance. We have run a multitude of diagnostic tests to determine of there was an error of implementation (we find no evidence of this) or if the sample who responded in the second study differed in significant ways from those in the first study (again finding no evidence). These surprising failures to replicate our earlier results suggest caution in interpreting the results from the first study.

Implications for practice

Asking multiple consents across several separate pages versus asking them together on one page does not seem to affect consent rates or other key outcomes. While quicker, asking a joint consent request (yes/no to all 5 domains) does not have significant advantages in terms of consent rates or understanding and confidence. Recall too that respondents in the qualitative interviews reported being overwhelmed when presented with multiple consent requests at once. The order in which a

series of linkage consent requests is made affects both overall rates of consent and consent rates to individual domains. While the effect is not yet clearly understood, and does not replicate across the two studies, it is clear that the order of the requests matters.

Future directions

Practitioners need to take care when asking for consent to multiple domains within a single survey. The order in which these are asked has potential consequences for the consent rates obtained to each of the domains. While we did not discern a clear pattern across the two studies that would lead to a specific recommendation, we urge caution in implementing multiple consent requests within a single survey. More research is needed to better understand the process of asking consent for multiple linkage domains. We did not test separating the consent requests across the survey (i.e., embedded within questions related to the domain in question), this is worth considering in future research, although it likely can only be done in the context of a longer survey.

5. Summary and outlook

The aim of this project was to improve our understanding about how to pursue the twin goals of maximizing consent and ensuring that consent is genuinely informed. Our main focus was on understanding how respondents process requests for data linkage: which factors influence their understanding of data linkage, which factors influence their decision to consent, and to open the black box of consent decisions to begin to understand *how* respondents make the decision. Focusing on the decision process, we also examined why web respondents are less likely to give consent than face-to-face respondents, and how best to ask for consent to multiple data linkages.

Using experimental data from the *Understanding Society* Innovation Panel and the PopulusLive online access panel, we have made progress: some of our findings have clear implications for best practice in asking respondents for consent to data linkage. Other findings suggest new hypotheses, with implications for survey design, that warrant further research.

The analyses reported in Section 4 examine different aspects of how respondents process consent requests. In this section we synthesise our findings to paint an overall picture of what we have learnt about how respondents make the consent decision and what remains to be tested. We also summarize the practical recommendations that result from our findings.

The probability that a respondent consents to data linkage is strongly associated with how well they understood the linkage process (as measured by knowledge test questions). However, this

association does not seem to be causal. When we experimentally manipulated the difficulty of the consent question wording, the easier version improved understanding of the linkage process, but did not increase consent. Conversely, when we asked the more difficult version of the consent request at the end of the questionnaire, respondents were less likely to consent than when we asked early on – although the position had no effect on their understanding of the request. This suggests that there are other variables that drive both understanding and willingness to consent.

There are several potential reasons why respondents might be less likely to consent to data linkage when asked late rather than early in the questionnaire. The qualitative interviews (Beninger et al. 2017) had suggested that this might be due to cognitive fatigue. However, since the position of the request had no effect on understanding, this hypothesis is not consistent with our findings. Another potential explanation is 'giving fatigue': respondents who have reached the end of the interview might feel they have provided enough information about themselves. However, if this were the case, we would expect the position to have an effect regardless of whether the question wording is easy or standard. A further potential explanation that is compatible with our findings is 'risk fatigue': respondents might be less willing to accept the uncertainty and potential risk of consenting when they are cognitively depleted. This hypothesis remains to be tested.

Respondents process the consent request in different ways: reflectively, considering the consequences of their decision, or less reflectively, based on gut feeling or habit. Those who process the request more reflectively spend more time on the question, are more likely to read additional materials, and consider information-based aspects of the request such as what information the government has about them and how much they know about the organisations involved. Those who base their decision on gut feeling or habit process the request more quickly, are less likely to read additional information, and more likely to consider what friends and family would do or wanting to avoid difficult decisions. Across the samples we interviewed, only 30 to 40 percent of respondents processed the request reflectively. In addition, respondents who answered the consent question online were less likely to process the request reflectively than those interviewed face-to-face. These findings suggest that simply providing additional information is unlikely to help increase informed consent.

Respondents who process the consent request more reflectively are more likely to consent and have better understanding of the request, than respondents who base their decision on gut feeling or habit. This may or may not be a causal relationship: the decision process might simply mediate the effects of characteristics of the respondent, the consent request and the survey design on consent outcomes. However, repeated interviews with respondents suggest that how they process the

consent request is not generally a stable trait. Respondents who make habit-based decisions are an exception: they are more likely to persistently base their decision on habit – and more likely to persistently decline consent. Further research is needed to test whether respondents can be moved to process consent requests more reflectively and whether doings so increases informed consent.

Being very confident in the consent decision is associated with *not* giving consent. That is, those who consent seem somewhat uncertain about their decision. This is no doubt because the decision is asymmetric: declining consent is risk free with certainty, whereas consenting involves unknown risks. Further research is needed to examine whether we can reduce the uncertainty respondents experience when making this decision, by matching the information provided with what respondents really want to know, and whether doing so increases informed consent.

Previous research has highlighted that situational factors might play an important role in the consent decision (Mostafa and Wiggins 2018). We hypothesized that how the respondent is feeling at the time of the survey might influence their decision: if they are feeling stressed or anxious they might be less willing to consent than if they were feeling happy and relaxed. However, our results show no such association between mood and consent. Our results however suggest that situational factors that invoke trust in the organisations involved in the linkage can increase consent.

The findings on decision processes and consent outcomes in face-to-face and web surveys point to further situational factors that might influence the consent decision. However, our findings are consistent with several competing hypotheses. Further experimental research is needed to disentangle these and identify which of the situational factors are most important. Such analyses would form the basis for thinking about practical implications of what can be done to increase informed consent in online surveys. Our results are consistent with social norms influencing the consent decision: if an interviewer is present, respondents seem to find it more difficult to decline the consent request than when they are answering the question on their own. Our results are also consistent with respondents valuing the possibility of asking the interviewer questions about the data linkage: even if audio-recordings of interviews suggest that they rarely take up this option, they might see the offer as part of the exchange. Our results are also consistent with the data collection technology priming respondents to be more or less concerned about the security of their data: respondents answering the survey online might feel unsure about where on the web their data are going, while face-to-face respondents feel that their data are handled in a secure environment.

The findings from this project have several practical implications for the design and implementation of consent questions in surveys:

- Writing consent questions in an accessible way – with low reading difficulty and including all relevant information in the question rather than in additional materials – can increase informed consent.
- Asking consent early in the questionnaire can increase consent, if the wording is difficult.
- Emphasizing trustworthiness of the organisations involved in the linkage (if appropriate), can increase consent.
- For multiple consent requests it doesn't seem to matter whether these are displayed individually on a sequence of pages, or jointly on a single page.
- When asking for multiple consents to data linkage, starting with domains for which consent rates are typically higher can increase consent rates.

In addition to these practical implications, our results point to new hypotheses about features of the linkage request that could potentially increase informed consent. These hypotheses remain to be tested:

- Reducing uncertainty about the consent decision, for example by matching the information provided about the linkage to what respondents would like to know and providing it in an accessible way, might increase informed consent. Future research in this area could be based on related research on consent forms for clinical trials.
- Manipulations that shift the respondent towards reflective processing of the consent request, for example by preceding the consent question with a notice that the respondent will be asked about their understanding of the request, might increase informed consent.
- In web surveys, highlighting the security of data transfers and that neither the survey nor the linked data are stored anywhere in the web might increase consent.
- Emphasising social desirability of giving consent might increase consent in web surveys.
- Offering an easy way of asking questions about the data linkage, for example via a chat function, might increase consent in web surveys.

A key conclusion from our research is that survey respondents process the request for data linkage consent in different ways. Acknowledging this, and tailoring the information provided to the needs of respondents may help increase informed consent. Simply providing more information on the linkage process – a common strategy employed by survey organisations and often advocated by ethics review boards – does not achieve the desired outcome. In contrast, our findings suggest that presenting the request in an easier-to-read format may increase understanding of the request, confidence in the decision, and consent rates. Minimising the gap in consent rates between self-

administered surveys completed online and those administered in-person by an interviewer remains a key challenge.

References

Al Baghal, T., Sloan, L., Jessop, C., Williams, M.L., and Burnap, P. (2019). "Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates across Three UK Studies." *Social Science Computer Review*, published online first.

Al-Shahi, R., Vousden, C., and Warlow, C. (2005). ''Bias from Requiring Explicit Consent from All Participants in Observational Research: Prospective, Population Based Study.'' *British Medical Journal* 331:942.

Bates, N. (2005). "Development and Testing of Informed Consent Questions to Link Survey Data With Administrative Records." in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 3786-3793.

Beninger, K., Digby, A., Dillon, G., and MacGregor, J. (2017). "How People Decide Whether to Give Consent to Link Their Administrative and Survey Data." *Understanding Society working paper 2017-13*, Colchester: University of Essex.

Benzeval, M., Bollinger, C.R., Burton, J., Couper, M.P., Crossley, T.F., and Jäckle, A. (2020). "Integrated Data: Research Potential and Data Quality," *Understanding Society working paper 2020-02*, Colchester: University of Essex.

Bryant, H., Robson, P.J., Ullman, R., Friedenreich, C., and Dawe, U. (2006). ''Population-Based Cohort Development in Alberta, Canada: A Feasibility Study.'' *Chronic Diseases in Canada* 27:51-59.

Buckley, B., Murphy, A.W., Byrne, M., and Glynn, L. (2007). ''Selection Bias Resulting from the Requirement for Prior Consent in Observational Research: A Community Cohort of People with Ischaemic Heart Disease.'' *Heart* 93:116-1120.

Calderwood, L. (2016). "Asking for Consent in a Mixed Mode Study – Issues and Outcomes." in *Presented at the Mixing Modes and Measurement Methods in Longitudinal Studies Workshop*. London: CLOSER.

Chaiken, S. (1980). "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39 (5): 752-766.

Cialdini, R.B., Vincent, J.E., Lewis, S.K., Catalan, J., Wheeler, D., and Darby, B.L. (1975). "Reciprocal Concessions Procedure for Inducing Compliance: The Door-in-the-face Technique." *Journal of Personality and Social Psychology* 31: 206-215.

Corcoran, K., and Mussweiler, T. (2010). “The Cognitive Miser’s Perspective: Social Comparison as a Heuristic in Self-judgements.” *European Review of Social Psychology* 21: 78-113.

Couper, M.P., Singer, E., Conrad, F.G., and Groves, R.M. (2008). “Risk of Disclosure, Perceptions of Risk, and Concerns about Privacy and Confidentiality as Factors in Survey Participation.” *Journal of Official Statistics* 24 (2): 255-275.

Couper, M.P., Singer, E., Conrad, F.G., and Groves, R.M. (2010). “Experimental Studies of Disclosure Risk, Disclosure Harm, Incentives, and Survey Participation.” *Journal of Official Statistics* 26(2): 287-300.

Das, M., and Couper, M.P. (2014). “Optimizing Opt-Out Consent for Record Linkage.” *Journal of Official Statistics* 30(3): 479-497.

Davis-Kean, P., Chambers, R.L., Davidson, L.L., Kleinert, C., Ren, Q. and S. Tang (2017) *Longitudinal Studies Strategic Review: 2017 Report to the Economic and Social Research Council,* Swindon: ESRC.

Dillard, J.P. (1991). “The Current Status of Research on Sequential-request Compliance Techniques.” *Personality and Social Psychology Bulletin* 17(3): 283-288.

Dolinski, D. (2011). “A Rock or a Hard Place: The Foot-in-the-Face Technique for Inducing Compliance Without Pressure.” *Journal of Applied Social Psychology* 41: 1514-1537.

Dunn, K.M., Jordan, K., Lacey, R.J., Shapley, M., and Jinks, C. (2003). ‘‘Patterns of Consent in Epidemiologic Research: Evidence from Over 25,000 Responders.’’ *American Journal of Epidemiology* 159: 1087-94.

Edwards, B., and Biddle, N. (in press). “Consent to data linkage: Experimental evidence on the impact of data linkage requests and understanding and risk perceptions.” In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley.

Eisnecker, P.S., and Kroh, M. (2016). “The Informed Consent to Record Linkage in Panel Studies: Optimal Starting Wave, Consent Refusals, and Subsequent Panel Attrition.” *Public Opinion Quarterly* 81(1): 131-143.

Fagerlin, A., Sepucha, K.R., Couper, M.P., Levin, C.A., Singer, E., and Zikmund-Fisher, B.J. (2010). “Patients’ Knowledge about 9 Common Health Conditions: The DECISIONS Survey.” *Medical Decision Making*, 30 (S1): 35S-52S.

Freedman, J.L., and Fraser, S.C. (1966). “Compliance Without Pressure: The Foot-in-the-door Technique.” *Journal of Personality and Social Psychology* 4: 195–202.

Fulton, J.A. (2012). *Respondent Consent to Use Administrative Data*. University of Maryland, unpublished PhD thesis. http://hdl.handle.net/1903/13601

Ghandour, L., Yasmine, R., and El-Kak, F. (2013). "Giving Consent without Getting Informed: A Cross-Cultural Issue in Research Ethics." *Journal of Empirical Research on Human Research Ethics* 8(3): 12-21.

Gigerenzer, G., and Gaissmaier, W. (2011). "Heuristic Decision Making." *Annual Review of Psychology* 62: 451-482.

Gigerenzer, G., and Goldstein, D.G. (1996). "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103: 650-669.

Groves, R.M. (2011) "Three Eras of Survey Research." *Public Opinion Quarterly*, 75(5): 861-71.

Harris, T., Cook, D.G., Victor, C., Beighton, C., Dewilde, S., and Carey, I. (2005). ''Linking Questionnaires to Primary Care Records: Factors Affecting Consent in Older People.'' *Journal of Epidemiology and Community Health* 59:336-38.

Hawley, S.T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., and Fagerlin, A. (2008). "The Impact of the Format of Graphical Presentation on Health-related Knowledge and Treatment Choices." *Patient Education and Counseling* 73(3):448-455

Huang, N., Shih, S.F., Chang, H.Y., and Chou, Y.J. (2007). ''Record Linkage Research and Informed Consent: Who Consents?'' *BMC Health Services Research* 7:18.

Jäckle, A., Beninger, K., Burton, J., and Couper, M.P. (in press). "Understanding Data Linkage Consent in Longitudinal Surveys." In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley.

Jenkins, S.P., Cappellari, L., Lynn, P., Jäckle, A., and Sala, E. (2006). "Patterns of Consent: Evidence from a General Household Survey." *Journal of the Royal Statistical Society, Series A* 169(4): 701-722.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kho, M.E., Duffett, M., Willison, D.J., Cook, D.J., and Brouwers, M.C. (2009). ''Written Informed Consent and Selection Bias in Observational Studies using Medical Records: A Systematic Review.'' *British Medical Journal* 338: b866.

Korbmacher, J.M., and Schroeder, M. (2013). "Consent When Linking Survey Data with Administrative Records: The Role of the Interviewer." *Survey Research Methods* 7 (2): 115-31.

Kreuter, F., Sakshaug, J.W., and Tourangeau, R. (2016). "The Framing of the Record Linkage Consent Question." *International Journal of Public Opinion Research* 28(1):142-52.

McNutt, L.-A., Waltermaurer, E., Bednarczyk, R.A., Carlson, B.E., Kotval, J., McCauley, J., Campbell, J.C., and Ford, D.E. (2008). "Are We Misjudging How Well Informed Consent Forms are Read?" *Journal of Empirical Research on Human Research Ethics* 3(1): 89-97.

Mostafa, T. (2016). "Variation within households in consent to link survey data to administrative records: evidence from the UK Millennium Cohort Study." *International Journal of Social Research Methodology* 19 (3): 355-375.

Mostafa, T, and Wiggins, R.D. (2018). "What influences respondents to behave consistently when asked to consent to health record linkage on repeat occasions?" *International Journal of Social Research Methodology* 21 (1): 119-134.

Olson, J.A. (1999). ''Linkages with Data from Social Security Administration Records in the Health and Retirement Study.'' *Social Security Bulletin* 62:73-85.

Pascale, J. (2011). "Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Participation (SHIPP)." *Survey Methodology working paper 2011-03.*, Washington, DC: Census Bureau.

Perrault, E.K., and McCullock, S.P. (2019). "Concise Consent Forms Appreciated—Still Not Comprehended: Applying Revised Common Rule Guidelines in Online Studies." *Journal of Empirical Research on Human Research Ethics* 14(4):299-306.

Petty, R.E., and Cacioppo, J.T. (1986). "The Elaboration Likelihood Model of Persuasion." *Advances in Experimental Social Psychology* 19: 123-205.

Ripley, K.R., Hance, M.A., Kerr, S.A., Brewer, L.E., and Conlon, K.E. (2018). "Uninformed Consent? The Effect of Participant Characteristics and Delivery Format on Informed Consent." *Ethics & Behavior* 28(7): 517-543.

Sakshaug, J.W., Couper, M.P., Ofstedal, M.B., and Weir, D.R. (2012). "Linking Survey and Administrative Records: Mechanisms of Consent." *Sociological Methods & Research 41(4):* 535–569.

Sakshaug, J.W., Hülle, S., Schmucker, A., and Liebig, S. (2017), "Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias." *Survey Research Methods* 11(2):171-188.

Sakshaug, J., and Kreuter, F. (2014). "The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey." *Public Opinion Quarterly* 78(1): 166-176.

Sakshaug, J.W., Schmucker, A., Kreuter, F., Couper, M.P., and Singer, E. (2019), "The Effect of Framing and Placement on Linkage Consent." *Public Opinion Quarterly* 83 (S1): 289-308.

Sakshaug, J.W., Stegmaier, J., Trappmann, M., and Kreuter, F. (2019). "Does Benefit Framing Improve Record Linkage Consent Rates? A Survey Experiment." *Survey Research Methods* 13(3): 289-304.

Sakshaug, J.W., Tutz, V., and Kreuter, F. (2013). "Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data." *Survey Research Methods* 7(2):133-44.

Sakshaug, J.W., Wolter, S., and Kreuter, F. (2015). "Obtaining Record Linkage Consent: Results from a Wording Experiment in Germany." *Survey Insights: Methods from the Field*. Retrieved from http://surveyinsights.org/?p=7288.

Sala, E., Burton, J., and Knies, G. (2012). "Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview and Interviewer Characteristics." *Sociological Methods and Research* 41(3): 414-39.

Sala, E., Knies, G., and Burton, J. (2014). "Propensity to Consent to Data Linkage: Experimental Evidence on the Role of Three Survey Design Features in a UK Longitudinal Panel." *International Journal of Social Research Methodology* 17(5):455-73.

Sepucha, K.R., Fagerlin, A., Couper, M.P., Levin, C.A., Singer, E., and Zikmund-Fisher, B.J. (2010). "How Does Feeling Informed Relate to Being Informed? The DECISIONS Survey." *Medical Decision Making*, 30 (S1): 77S-84S.

Schwarz, N., and Sudman, S. (Eds.) (1992). *Context Effects in Social and Psychological Research*. New York: Springer-Verlag.

Singer, E., and Couper, M.P. (2010). "Communicating Disclosure Risk in Informed Consent Statements." *Journal of Empirical Research on Human Research Ethics* 5(3): 1-8.

Strack, F., and Deutsch, R. (2004). "Reflective and Impulsive Determinants of Social Behavior." *Personality and Social Psychology Review* 8: 220-247.

Stunkel, L., Benson, M., McLellan, L., Sinaii, N., Bedarida, G., Emanuel, E., and Grady, C. (2010). "Comprehension and Informed Consent: Assessing the Effect of a Short Consent Form." *IRB* 32(4): 1-9.

Tait, A.R., Voepel-Lewis, T., Nair, V.N., Narisetty, N.N., and Fagerlin, A. (2013). “Informing the Uninformed: Optimizing the Consent Message Using a Fractional Factorial Design.” *JAMA Pediatrics* 167(7):640-46.

The Nuffield Foundation (2016). "Building Bigger and Better Data for the Public Good: Administrative Data for Survey Design and Linkage." Report from a Collaborative Workshop Hosted by the Nuffield Foundation, CLOSER, ESRC and RSS: The Nuffield Foundation.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K., and Gaia, A. (2018). “Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a Large-scale Longitudinal Study in the UK.” *Survey Methods: Insights from the Field.* Retrieved from https://surveyinsights.org/?p=9734.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.

Tversky, A., and Kahneman, D. (1974). “Judgment under Uncertainty: Heuristics and Biases.” *Science* 185: 1124-1130.

Yawn, B.P., Yawn, R.A., Geier, G.R., Xie, Z., and Jacobsen, S.J. (1998). ‘‘The Impact of Requiring Patient Authorization for Use of Data in Medical Records Research.’’ *Journal of Family Practice* 47:361-65.

University of Essex. Institute for Social and Economic Research (2019) *Understanding Society*: Innovation Panel, Waves 1-11, 2008-2018. [data collection]. 10th Edition. UK Data Service. SN: 6849 http://doi.org/10.5255/UKDA-SN-6849-11.

Weir, D.R., Faul, J.D., and Ofstedal, M.B. (2014). “The Power of Persistence: Repeated Consent Requests for Administrative Record Linkage and DNA in the Health and Retirement Study.” Paper presented at the *Panel Survey Methods Workshop*. Ann Arbor, MI.

Weiß, B., Beuthner, C., Silber, H., Keusch, F., Menold, N., and Schröder, J. (2019). “Consenting to Data Linkage: The Role of Data Domain, Framing, Device, Incentives, and Respondent Characteristics.” Presentation at *Mobile Apps and Sensors in Surveys (University of Mannheim and Utrecht University)*.

6. Appendix Tables

Appendix Table 1: Predictors of consent, by sample

	IP	IP	AP1	AP1	AP2	AP2
Female	0.017	0.008	-0.081***	-0.081***	-0.101***	-0.101***
[Omitted: age 16-40]	.	.	.	.	.	.
41-59	-0.069*	-0.068*	0.012	0.013	-0.009	-0.01
60+	0.021	0.019	0.082**	0.085**	0.061	0.06
[Omitted: degree]	.	.	.	.	.	.
A/AS level	0.043	0.044	0.023	0.023	0.031	0.032
GCSE or lower	0.01	-0.001	-0.001	-0.002	0.049	0.049
Reads daily	0.035	0.01	-0.005	-0.006	0.023	0.021
In work	0.029	0.031	-0.028	-0.026	-0.062*	-0.062*
Household size	-0.005	0.002	-0.006	-0.006	-0.008	-0.009
[Omitted: home owned outright]	.	.	.	.	.	.
Home owned with mortgage	-0.034	-0.027	-0.007	-0.006	0.131***	0.131***
Home rented or other	0.081**	0.043	-0.027	-0.022	0.074*	0.072*
Worried about privacy	-0.052*	-0.037	-0.073**	-0.072**	-0.04	-0.04
Concerned about data security	-0.071**	-0.066**	-0.05	-0.049	-0.043	-0.042
Attitudes towards data sharing	-0.032***	-0.027***	-0.034***	-0.034***	-0.026***	-0.026***
No. of online data sharing behaviours	0.012*	0.014**	0.012*	0.011	0.009	0.009
HMRC data are (highly) sensitive	-0.044*	-0.039*	-0.066**	-0.067**	-0.05	-0.051
Trusts HMRC somewhat/a lot	0.221***	0.224***	0.194***	0.194***	0.140***	0.138***
Knows what data HMRC have	0.068**	0.023	0.056**	0.055**	0.090***	0.090***
Survey completed online	.	-0.230***	.	.	.	.
Easy consent question wording	.	0.034	.	0.031	.	.
Late position in questionnaire	.	-0.069**	.		.	.
Easy wording x face-to-face survey	.	0.046	.	.	.	.

Appendix Table 1: Predictors of consent, by sample - continued

	IP	IP	AP1	AP1	AP2	AP2
Easy wording x web survey	.	0.017	.	.	.	.
Easy wording x early position	.	-0.063	.	.	.	.
Easy wording x late position	.	0.065***	.	.	.	.
More information with follow up	.	.	.	-0.024	.	.
More information no follow-up	.	.	.	-0.092**	.	.
Default consent question wording	.	.	.	-0.029	.	.
Trust priming	.	.	.	.	.	0.044*
N	2661	2661	2563	2563	1921	1921

Notes: Average Marginal Effects from logit models.

Appendix Table 2: Predictors of objective understanding of the linkage request, by sample

	IP	IP	AP1	AP1	AP2	AP2
Female	0.032	0.009	-0.016	-0.018	-0.066	-0.068
[Omitted: age 16-40]	.	.	.	.	.	.
41-59	0.103	0.097	0.093	0.093	0.208*	0.207*
60+	0.086	0.103	0.081	0.081	0.209	0.208
[Omitted: degree]	.	.	.	.	.	.
A/AS level	-0.210*	-0.225*	-0.055	-0.053	-0.114	-0.112
GCSE or lower	-0.521***	-0.549***	-0.472***	-0.471***	-0.365***	-0.365***
Reads daily	0.07	0.006	-0.052	-0.043	-0.053	-0.056
In work	0.006	0.02	-0.116	-0.116	-0.165*	-0.165*
Household size	-0.067*	-0.043	-0.019	-0.02	-0.003	-0.004
[Omitted: home owned outright]	.	.	.	.	.	.
Home owned with mortgage	0.051	0.063	-0.128	-0.115	0.071	0.072
Home rented or other	0.018	-0.101	-0.128	-0.131	-0.156	-0.158
Worried about privacy	-0.095	-0.063	-0.320***	-0.320***	-0.214*	-0.215*
Concerned about data security	-0.137	-0.115	0.238*	0.246**	0.151	0.153
Attitudes towards data sharing	0.02	0.033*	0.002	0.002	0.016	0.016
No. of online data sharing behaviours	0.125***	0.132***	0.046*	0.047**	0.047*	0.047*
HMRC data are (highly) sensitive	0.023	0.026	0.240**	0.233**	0.038	0.037
Trusts HMRC somewhat/a lot	0.419***	0.414***	0.304***	0.312***	0.309***	0.306***
Knows what data HMRC have	0.492***	0.382***	0.126	0.113	0.261***	0.261***
Survey completed online	.	-0.712***	.	.	.	.
Easy consent question wording	.	0.413***	.	0.362***	.	.
Late position in questionnaire	.	0.028	.	.	.	.
Easy wording x web survey	.	-0.037	.	.	.	.
Easy wording x late position	.	0.192	.	.	.	.
More information with follow up	.	.	.	-0.079	.	.
More information no follow-up	.	.	.	-0.047	.	.

Appendix Table 2: Predictors of objective understanding of the linkage request, by sample - continued

	IP	IP	AP1	AP1	AP2	AP2
Default consent question wording	.	.	.	0.034	.	.
Trust priming	.	.	.	.	.	0.074
Constant	3.533***	3.501***	4.239***	3.953***	4.156***	4.128***
N	2661	2661	2563	2563	1921	1921

Notes: Coefficients from OLS models.

7. Appendix: Main survey questions used for analyses[3]

Respondent background characteristics

Privacy concerns

In general, how worried are you about your personal privacy?

1. Very worried
2. Somewhat worried
3. Not very worried
4. Not worried at all

Data security concerns

Different private and public organizations have personal information about us. How concerned are you about whether or not they keep this information confidential?

1. Very concerned
2. Somewhat concerned
3. Not very concerned
4. Not concerned at all

Consent questions

Standard consent question wording

We would like to add records held by HM Revenue and Customs, or HMRC, containing information on your employment and self-employment history, your income, National Insurance contributions and tax credits. All information will be used for research purposes only by academic or policy researchers under restricted access arrangements which make sure that the information is used responsibly and safely.

Please read this leaflet and look at this diagram [Version B] explaining how we would like to attach your HMRC records to the answers you have given in this study.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1. I have read the leaflet and am happy to give consent
2. I do not want to give consent

---

[3] The full Innovation Panel and Access Panel questionnaires are available at: https://www.iser.essex.ac.uk/research/projects/understanding-and-improving-data-linkage-consent-in-surveys.

Easy consent question wording

We would like to add records held by HM Revenue and Customs, or HMRC, to the answers you have given in this study. If you agree:

- We will send HMRC your name, address, sex and date of birth so that they can identify the records they have about you. The HMRC records contain information about your current and previous employment, your income, National Insurance contributions and tax credits.
- We will not send HMRC the answers you have given in this study.
- HMRC will send us your records. These will contain an anonymous identification number but not your name, address, sex or date of birth.
- We will add the HMRC records to the answers you have given in this study.
- We will make the combined anonymous data available for academic and policy research purposes only.
- Access to the data will be restricted and controlled, to make sure that researchers use the information responsibly and safely.
- This will not affect the way that you deal with the HMRC in any way.

Please read this leaflet and look at this diagram [Version A] for further information.

Do you give permission for us to pass your name, address, sex and date of birth to HMRC for this purpose?

1. Yes
2. No

Consent wording: trust experiment

The next question is about linking the information you provide in this survey, to data that HM Revenue and Customs, or HMRC, hold about you.

[IF trust treatment group: HMRC is a trusted data holder [display trust symbol]]

1. Continue

Follow-up questions about consent decision process

Subjective understanding of consent request

How well do you think you understand what would happen with your data, if you allowed us to link it to records held by HM Revenue and Customs?

1. I do not understand at all
2. I understand somewhat
3. I mostly understand
4. I completely understand

Confidence in linkage consent decision

We are interested in how people decide whether or not to give us permission to add data held by HM Revenue and Customs to the answers they have given in this study.

How confident are you about the decision decisions you made?

1. Very confident in my decision
2. Confident in my decision
3. Somewhat confident in my decision
4. Not confident in my decision

Consent decision process

How did you decide whether to say “yes” or “no” in response to the question about data linkage?

Please select all of the answers that apply to you.

1. I thought about what would happen if I said “yes” or “no”
2. Instinct or gut feeling
3. I said what I usually say when I’m asked for information that is very personal
4. [IF Access Panel sample: 4 I thought about how much I trust the organisations involved]
5. Something else (please specify)

Cognitive effort

On a scale of 0-10, where 0 is no effort at all and 10 is a great deal of effort, how much effort did you put into coming up with your answer about data linkage?

Please select one only

0. No effort at all
…
10. A great deal of effort

Factors considered when making consent decision

Scripting note: randomise items, display as grid with radio buttons. Five point response scale with end labels but no numbers ("1Played no role", 2, 3, 4, 5 "Played very big role")

When you were deciding whether or not to allow your data to be linked, how much of a role did each of the following aspects play in your decision?

- What information the government has about me
- The way I usually make decisions
- My gut feelings
- How much I trust the organisations involved
- What friends and family would do
- Avoiding difficult decisions
- Wanting to be helpful
- The benefits to society
- How much information I've already shared
- How much I know about the organisations involved
- Who might have access to my information
- Whether the Government would use this data to check on me
- Recent news stories related to data security
- The chances of my personal data getting into the wrong hands or being misused

Objective understanding of data linkage

To help us understand whether the explanation we gave you about linking HMRC data and your answers to this study was clear or unclear, here are a few statements about how the linkage is done. Please specify whether you think each of the statements is true or false.

Answer categories: True/false for each row

- Every researcher can access the combined data via the Internet [false]
- HM Revenue and Customs will combine the information they have with your answers to this study [false]
- Researchers using the data will only have access to anonymous data [true]
- The combined data can be used by HM Revenue and Customs to check that you have been paying your taxes [false]
- HM Revenue and Customs will send us the information they have about you [true]
- Your name, address, sex, and date of birth will be saved with the linked data [false]
- We will send your name, address, sex, and date of birth to HM Revenue and Customs [true]
- HM Revenue and Customs will send us future data about you, unless you object in writing [true]