Understanding Society Working Paper Series

No. 2021 – 05

December 2021

How Do Survey Respondents Decide Whether to Consent to Data Linkage?*

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

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

[2] University of Michigan, [3]European University Institute, [4]University of Konstanz

*Author order is alphabetical. [#] Corresponding Author.

Understanding Society

UKRI Economic and Social Research Council

Non-technical summary

Survey data are increasingly being linked to administrative records to maximize the value of the data for social and health research while minimizing respondent burden. Obtaining explicit consent for such linkage from survey respondents and ensuring that such consent is informed are important for ethical and legal reasons. While a growing body of research has examined the correlates of consent to record linkage and explored ways to increase consent rates, we still know very little about how people make a decision to consent to record linkage or not. In this paper we use experimental data collected in five different surveys – including face-to-face and online surveys from the *Understanding Society* Innovation panel and three different online surveys administered to members of an opt-in access panel – to explore the process of consent decision-making.

A key finding is that respondents report using a variety of different decision processes, some of which are more "reflective" (for example, considering the consequences of consent or their trust in the organisations involved) while others suggest the use of less or different information (for example, based on "gut-feeling"). Importantly, across all samples, less than 40% of respondents report using a reflective decision process exclusively. These self-reported processes are corroborated by markers of effort: those who report a more reflective processing of the request take longer to respond, are more likely to read additional information about the linkage and are more likely to self-report higher levels of effort in answering the question. Reported decision processes align with the nature and quantity of information respondents report drawing on in making the decision. We find that the self-reported decision process is weakly predicted by background characteristics and not affected by several survey design manipulations.

Finally, we find that self-reported decision processes are associated with consent outcomes. More reflective decision processes are associated with higher consent, greater comprehension and greater confidence in the decision. Conversely, decisions described as "gut" or habitual are associated with lower consent and lower comprehension.

These findings have important implications for practice. Many attempts to achieve higher consent rates and more informed consent involve the provision of additional information, but these have had limited success. Our analyses point to a possible explanation: if many of those who withhold consent are using a very rapid and less reflective decision process, additional information is unlikely to be incorporated into their 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 that weighs the pros and cons of the decision, or one based on trust in the relevant organisations. Our results indicate that this could have a double benefit: in our data, those employing the more reflective decision processes are both more likely to consent and have greater understanding of the request. That is, they are more likely to give informed consent.

How Do Survey Respondents Decide Whether to Consent to Data Linkage?

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

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

[2] University of Michigan, [3]European University Institute, [4]University of Konstanz

Abstract: Linkages between surveys and administrative data provide an important opportunity for social and health research, but such linkages often require the informed consent of respondents. We use experimental data collection across five different samples to study how consent decisions are made. Only about a third of respondents report using a "reflective" decision process, considering the consequences of consent or their trust in the relevant organisations. Many more use an "instinctive" process (such as "gut feeling"). More reflective decision processes are associated with higher rates of consent, greater comprehension of the proposed data linkage and greater confidence in the decision.

Keywords: record linkage, informed consent.

JEL classification: C81, C83, D91

Acknowledgements: Author order is alphabetical. This research was funded by the Nuffield Foundation (www.nuffieldfoundation.org) with co-funding from the Economic and Social Research Council (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 views expressed here are those of the authors and not necessarily of the Nuffield Foundation or the ESRC. We thank Brendan Read for excellent research assistance and are 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, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom, aejack@essex.ac.uk.

1. Introduction

The complementary strengths of survey and naturally occurring data (administrative data or transaction data) mean that "integrated" data that combines the two provides a particularly valuable basis for future empirical work in the social and human sciences (Groves 2011; Meyer et al., 2015; Davis-Kean et al. 2017; Benzeval et al. 2020). For legal and ethical reasons, linking administrative or other naturally occurring data with survey responses often requires respondent consent to data linkage. Failure to obtain consent leads to reduced rates of linkage and so to reduced samples and potential representation errors (selection biases). Survey managers and data producers therefore wish to maximize consent rates, but the same legal and ethical considerations imply that the appropriate objective is to maximize *informed* consent.

The existing empirical evidence on survey respondents' consent to data linkage contains a number of puzzling findings. Empirical correlates of consent are inconsistent (e.g., Peycheva et al., 2021; Sala et al. 2012). Individual survey respondents appear to have a latent "willingness to consent" in multiple consent requests asked within one interview (Jenkins et al., 2006; Mostafa, 2016; Walzenbach et al., 2021) but evidence of a latent willingness to consent over time is weaker (Mostafa and Wiggins, 2018) and many respondents who decline to consent reverse this decision if asked again at a later date (Weir et al. 2014; Jäckle et al. 2021a). Efforts to increase consent rates through experimental manipulation of the requests have produced mixed results (Kreuter et al. 2016; Sakshaug et al. 2013; Sala et al. 2014). Perhaps more worrying, comprehension of linkage requests appears to be poor (Das and Couper 2014; Edwards and Biddle, 2021; but see Sakshaug et al. 2021). Overall, the process by which a particular survey respondent in a particular context does or does not provide consent is poorly understood.

Our goal in this paper is to examine *how* individuals make consent decisions and – more explicitly – how *informed* consent happens. The consent decision is necessarily made within a limited time frame (a survey or interview), and with incomplete information. We hypothesize that survey respondents predominantly use heuristic decision processes to make the consent decision (and so we avoid the common systematic/unsystematic labels for decisions), but that these processes are heterogeneous across individuals and contexts. Heuristic decision processes differ in the amount and nature of the information that is used in making the decision; we refer to more information-intensive heuristics as "reflective". Finally, we also hypothesize

that more reflective decision processes are associated with higher consent propensities and greater comprehension.

To explore these hypotheses, we implemented experimental data collection across five different survey samples with different characteristics: face-to-face and online, population and convenience, cross-section and panel. Respondents were presented with a consent request and additional information was then collected both from survey paradata and follow-up questions, including self-reports of how the consent decision was made.

Using these data, we first document the extent to which self-reports indicate heterogeneity in the consent decision-making process and we corroborate these self-reports with more objective measures, such as time taken to respond to the consent request. Next, we document how the decision process employed to respond to a consent request is affected by background characteristics of the respondent and by features of the survey context. Three of the latter (the placement of the consent request, the readability of the consent request, and whether respondents were "primed" to reflect on their trust in the survey organisation), were experimentally manipulated in at least one of our samples. Finally, we examine the associations between self-reported decision process and consent outcomes: the decision itself, measures of comprehension of the linkage process, and a measure of subjective confidence in the decision. Two of our samples are linked longitudinally, allowing us to examine the longitudinal association between decision process and consent as well.

2. Background, Conceptual Framework, Hypothesis

Our view is that the decision to consent to data linkage is not an overly consequential one for the individual and is based on relatively limited information presented in a constrained time frame. Given this, the decision is likely to be made on the spot, with participants making a relatively quick decision based on a limited amount of information presented, and using information external to the request itself, such as knowledge of or trust in the organisations involved. The decision process is heuristic but may involve varying degrees of information and reflection.

Our conceptual thinking and the research design for this paper were informed by qualitative in-depth interviews, in which *Understanding Society* respondents were asked how they came up with their decision to consent or not to consent to

administrative record linkage (Beninger et al. 2017; Jäckle et al. 2021a). In interpreting the findings from the qualitative interviews, we drew on several strands of literature on decision making.

There is a large literature on dual-process theory or two-system processing, covering a wide range of research areas from medical decision making to behavioural economics. 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 Cacioppo 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; see also Gigerenzer 2000).

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." Similarly, Galotti (2007) found that non-experts making important real-life decisions "consistently constrained the amount of information they considered," consistent with the notion of bounded rationality (see Simon 1957, 1959) and suggesting adaptive strategies even for complex and consequential decisions.

Where does the request for consent to data linkage fit into this literature? It 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 largely unknown; much of the information provided in surveys is about the *process* of linkage, rather than the

consequences). While consent materials are often written to provide as much information as possible about the request, assuming a careful evaluation of the pros and cons, consent decisions seem to be made relatively quickly (Desch et al. 2011; Ghandour, Yasmine and El-Kak, 2013; Jäckle et al. 2021d; McNutt et al. 2007) and change over time (Weir et al. 2014; Jäckle et al. 2021a). Further, the decision is an unbalanced yes/no choice. A "yes" decision brings potential risks but few tangible benefits for the respondent, while there are few, if any, negative consequences of a "no" decision for the individual. Contrast this, for example, with a choice between surgery and medication in the medical decision-making literature.

The decision to consent (or not) to record linkage is also made in the context of a survey. For most survey questions, where the stakes are relatively low and the consequences of an "incorrect" answer are negligible, respondents may engage in "satisficing" (Simon 1957; Krosnick 1991; Krosnick and Alwin 1987) or taking cognitive shortcuts. Switching to a more deliberative approach to consider the pros and cons of consenting to record linkage may run counter to the default mode of responding in surveys, where respondents are sometimes encouraged to give the first answer that comes to mind.

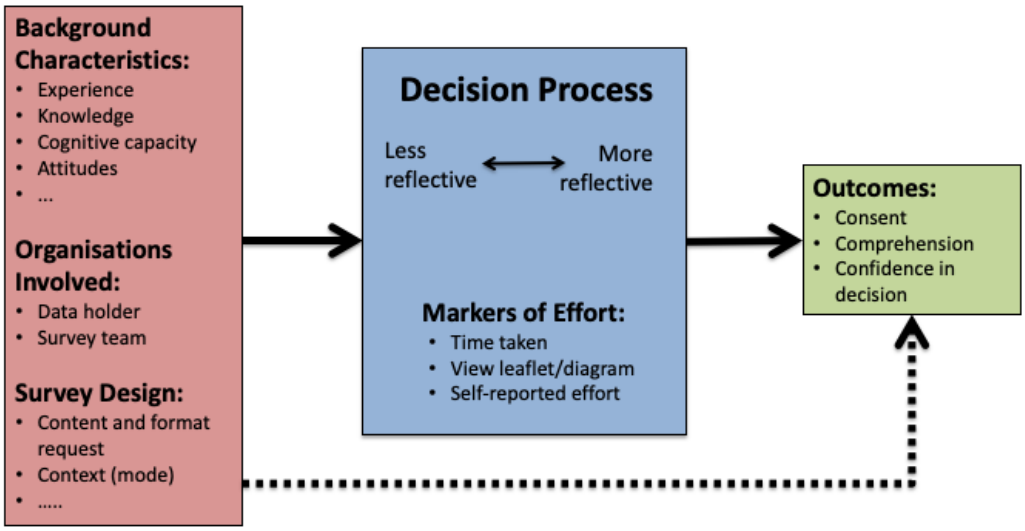
Thus, given the constraints of time and information, and the low stakes, we hypothesize that respondents use a variety of strategies to deal with the consent request. However, we also hypothesize that the selected decision process varies across individuals and contexts, differing in the amount and nature of the information that is considered. Some individuals will attempt to consider the costs and benefits of consent, or trust in the survey organisation or data holder. We will refer to such decisions as "more reflective" (though this should not be taken to imply fully rational decision making). Others may be based on habit or gut feeling. We refer to these as "less reflective." That is, there is variation in how information intensive the decision process is, and the extent to which central arguments (pros and cons of consent) versus peripheral cues (heuristics) are used in the process of making a decision.

There are many factors that could affect the processes survey respondents use when making a decision about consent. These are illustrated in Figure 1. 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 organisations involved, including both the survey organisation and the data holder. Participants' knowledge of, experience with, and trust in, those organisations 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. Much of the experimental research on consent has focussed on manipulating the content and format of the request (e.g. gain versus loss framing, longer versus shorter descriptions of the linkage process, etc.). This literature, which is reviewed in Jäckle et al. (2021d), has found small and often contradictory effects on consent.

In the conceptual framework described in Figure 1, the decision process mediates the effect of these background variables on outcomes, including the decision itself and comprehension of the request. At the same time, we do not rule out that characteristics of the individual, survey organisation, consent request and survey context may have direct effects on these outcomes (as indicated by the dotted arrow in Figure 1). An important implication of this is that it is very challenging to estimate the causal effect of the decision process on outcomes. To do so requires a source of variation (e.g., an instrumental variable) that affects outcomes *only* through the decision process, or a strong sequential ignorability assumption (Imai et al. 2011; Keele et al. forthcoming). It is not clear how either requirement could be met. Nevertheless, associations between decision processes and consent outcomes are suggestive, and to our knowledge, have not previously been documented. One of our samples also allows us to look at longitudinal associations between decision processes and consent outcomes. We can however plausibly estimate the (full) causal effects of pre-determined characteristics of individuals on decision processes, as well as the causal effects of experimentally-manipulated survey design features. We now turn to describing our data.

Figure 1: Conceptual Framework



3. Data, Measures and Experimental Design

3.1 Samples and Research Stages

We collected five samples from two studies in three stages. In the first stage we collected data in Wave 11 of *Understanding Society*'s Innovation Panel. *Understanding Society* is also known as the UK Household Longitudinal Study. The Innovation Panel is a probability sample of households in Great Britain that is used for methodological testing and experimentation and its design mirrors 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. Wave 11 was fielded in May to October 2018 by Kantar Public and NatCen Social Research (University of Essex, 2021). In the text below we refer to this survey as IP11.[1] Where we present results separately for

[1] 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.

face-to-face and web respondents, we refer to IP11A and IP11B, respectively, but our focus is on the replication of findings across samples with different characteristics.[2]

Since the sample size constrained the number of experimental treatment groups we could implement, we fielded additional surveys using an access panel. The PopulusLive access panel (AP) is a non-probability online panel in the UK with around 130,000 active sample members at that time, who are recruited through web advertising, word of mouth, and database partners. To enable some comparison with the Innovation Panel sample, the sample was restricted to Great Britain and quotas based on age, gender and education were set to match the characteristics of the IP11 sample.

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 respectively. 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 AP samples included other experiments not reported on here (see Jäckle et al. 2021b,d and Walzenbach et al. 2021). Table 1 presents a brief summary of the key features of the five samples used in our analyses. Additional information on response and participation rates is presented in Appendix A.

Table 1: of Samples and Experiments

	IP11A	IP11B	AP1.1	AP1.2	AP2
Sample type	prob	prob	non-prob	non-prob	non-prob
Mode	FTF	Web	Web	Web	Web
Analysis sample size	1363	1299	1034	816	967
Experiments					
Wording (easy/standard)	x	x	x	x	
Location (early/late)	x				
Trust prime (yes/no)					x

3.2 Measures

a. Outcomes: Consent, Understanding, Confidence

The key outcome we examine is consent to linkage to income, employment and tax records held by the UK tax authority, HM Revenue and Customs (HMRC). The full

[2] For further details of the mode experiment, see Jäckle et al. (2021c). In the current paper, we define samples by mode of *response*, rather than assigned mode.

wording of the question is reproduced in Appendix B. Item nonresponse rates for the consent request were low, ranging from 0 in the AP1 sample to 3.2% in the IP11A sample. In practice, failure to answer a consent question means that consent has not been granted, so for this question we code item nonresponse as non-consent.

Objective understanding was measured using a series of eight true/false questions about the data linkage process (see Appendix B4). These are based on a similar test of understanding by Das and Couper (2014). Across samples less than 4% of respondents answered some but not all eight test questions. We assume that question-specific nonresponse indicates an inability to answer that question, and so don't know and refusal answers are coded as incorrect answer. In contrast, the test score is set to missing for respondents who did not answer *any* of the eight test questions. This was below 1.5% in all samples except for IP11B, where it was 11% of the sample.[3]

Subjective understanding was measured with a single item asking *"How well do you think you understand what would happen with your data …,"* with a four-point response scale ranging from *"I do not understand at all" to "I understand completely."* Confidence was similarly measured with a four-point response scale ranging from *"Very confident in my decision"* to *"Not confident in my decision"*. Full details of these questions are presented in the Appendices A5 and A6. For both measures, item nonresponse was below 2% in all samples.

b. Self-reported Decision Process

Based on responses in earlier qualitative interviews (see Beninger et al. 2017 and Jäckle et al. 2021b), we drafted a question to elicit the process of decision-making as follows:

"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.

- *I thought about what would happen if I said "yes" or "no"*
- *Instinct or gut feeling*
- *I said what I usually say when I'm asked for information that is very personal*

[3] We have not been able to pinpoint the cause of the significantly higher item nonresponse for this item in this sample, despite extensive investigation. Sensitivity analyses making different assumptions about these missing cases does not alter the key findings.

- *I thought about how much I trust the organisations involved (AP1.1. and AP1.2 only)*
- *Something else (please specify)"*

After reviewing and coding the open text responses in the "something else" category from the AP pilot survey, we decided to add the response option about trust in the organisations involved to the AP1.1 and AP1.2 surveys. Nonresponse to this question – defined as not selecting any option, not even "other" – was completely absent in the AP samples, and 1.3% and 3.6% in the IP11A and IP11B samples respectively.

Note that respondents could select more than one option. For the analyses below, we employ two coding schemes for the responses to this question. For descriptive statistics and testing for differences between groups defined by response to this question, we adopt an "exclusive" coding system, where we examine the five most common combinations of one or more response options. These patterns define mutually exclusive groups of respondents, and they cover the majority of respondents (between 72% and 87%, depending on the sample). However, when we estimate models in which the self-reported decision process is either the dependent or independent variable, we employ an "inclusive" coding. For example, we code "trust" as one if the respondent selected that option, either alone or in conjunction with other options (and zero otherwise). This allows the full sample to be employed in model estimation and means that the contrast is between all those who report trust, and all those who do not.

c. Markers of Effort

The time taken to answer the consent question (in seconds) was derived from keystroke paradata. This measure is right-skewed, so we performed a logistic transformation for analysis.

For online respondents, we also used paradata to identify whether they clicked on the links to access additional information (leaflet and diagram) about the data linkage. For face-to-face respondents in the IP, we used an interviewer observation asking whether the respondent read the leaflet or whether the interviewer explained the diagram to the respondent.

In samples AP1.2 and AP2 we added a subjective measure of effort, asking respondents *"On a scale of 0 to10, 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?". There was no item nonresponse on this question in either sample.

Further, in the AP2 sample we also asked people what role alternative factors played in making their decision. The full list of possible factors is given in Table 3 below and Appendix B7, but included options such as *"what information the government has about me"*, *"how much I know about the organisations involved"*, and *"the benefits to society".* The order of these response options was randomized to reduce primacy effects. Each option was rated on a five-point scale with 1 being *"played no role"* and 5 being *"played a very big role"*. There was no complete nonresponse for this item battery, but 0.5% of respondents had missing values on some of the factors.

d. Model covariates

Finally, we also include selected socio-demographic background covariates in our multivariable models: sex, age, education, and employment status.[4] Among these covariates, item missingness is very low: below 1% for all variables in all samples except education, which reaches 1.9% in IP11B and 3.1% in IP11A.

Given the very low rates of item nonresponse throughout, we code nonresponse as described in this section and conduct complete-case analyses below.

3.3 Survey Experiments

a. Question wording (IP11A, IP11B, AP1.1 and AP1.2):

This experiment was designed to manipulate the difficulty of comprehending the consent request. It was replicated in four of our samples. In each sample, half of respondents were randomly allocated to the 'standard' wording of the consent question, 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. Both versions offered a leaflet and diagram, but the latter was enhanced for the easy version. The wording of the response options was also simplified for the easy version. Further details can be found in Appendix B.

b. Question placement (IP11A)

[4] We considered a number of additional predictors including household size, home ownership, measures of mood, and, in the IP samples, time in panel (a measure of commitment to the panel). As these did not have statistically significant effects in any sample, they are dropped from the final models reported below.

This experiment was implemented only in IP11A. Half the sample was asked for consent early in the interview (after a series of socio-demographic questions); the other half was asked at the end of the survey (before the self-completion CASI module).

c. Trust priming (AP2)

This experiment was implemented only in AP2. Half of the sample (the no priming condition) was shown an introductory screen saying, *"The next question is about linking the information you provide in this survey, to data that HMRC hold about you."* Those in the trust priming condition saw an additional statement on this screen: *"HMRC is a trusted data holder,"* followed by an icon symbolizing data security.

4. Results

4.1 Self-Reported Decision Process

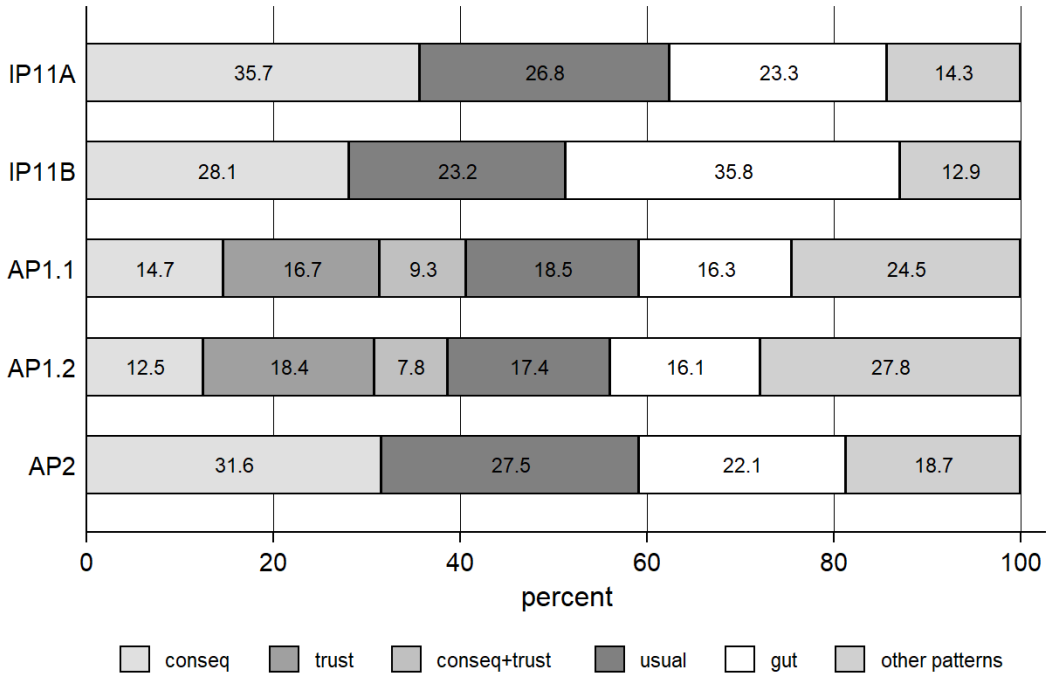
When asked after the fact how they made their decision to grant or deny consent to data linkage, the majority of respondents chose a single response option (e.g. consequentialist thinking or gut feeling). Multiple answers were allowed, but relatively infrequent; each combination was reported by less than 3% of respondents, with one exception: In the AP1.1 and AP1.2 samples, where, as noted earlier, we additionally offered "trust" as a response option, a substantial fraction of respondents (9.3 and 7.8% respectively) selected both consequentialist and trust response options.

In light of this, Figure 2 uses an exclusive coding of the decision process and shows the five most common patterns, subsuming all other combinations of answers (as well as all those who selected "other") in "other patterns". For example, the label "consequentialist" refers to respondents who exclusively selected the option "I thought about what would happen if I said 'yes' or 'no'". Across samples, this option was chosen (alone) by 12.5-35.7% of respondents. We consider both "consequentialist" and "trust" to be more reflective decision processes. They are often reported together, and indicate reflection, either on the consequences of consent, or on the organisations involved.

Other respondents reported decision processes that were less reflective or more instinctive. For example, across all samples, 17 to 28% of respondents selected only the option "*I said what I usually say*" (labelled "usual" in Figure 2) and 16 to 34% reported only "instinct or gut feeling" (labelled "gut" in Figure 2).

It is clear that respondents make decisions in heterogeneous ways. A significant fraction of respondents report thinking about the consequences of consent, but another group instead consider their trust in the survey organisation and data holder, and further groups make less reflective decisions, relying on a gut feeling or instinct, or on habit or usual practice.

Figure 2: Frequencies of Self-Reported Decision Processes



In AP1.2, respondents who chose several decision processes were asked to specify the most important one in a follow-up question. Of these, 36.8% reported trust to be their prime motivation behind their consent decision, followed by 24.7% who picked a consequentialist approach, 18.9% who did the usual, 16.7% who relied on their gut feeling and 3% who picked other. If we only consider one most important decision process per respondent, trust remains the single most important decision process in the AP1.2 sample. Trust was either the only or main decision process reported by 30% of all AP1.2 respondents.

Importantly, we find that the fraction of respondents reporting an *exclusively* reflective decision process (consequentialist and/or trust) is 40 percent or less in all samples (ranging from 27 percent in IP11A to 40 percent in AP1.1). In contrast, the fraction of respondents who exclusively report unreflective decision processes ("gut"

or “usual”) range from 37 percent in the AP1 samples to 60 percent in IP11A. Taken on their face, these self-reports are an important challenge to the usual strategy of increasing consent through the provision of additional information.

Are these self-reports of decision process meaningful? To assess this question, we looked at markers of decision effort. These are reported in Table 2 and they are consistent with the self-reports.

Starting with response time, in AP1.1 respondents who reported “consequentialist”- or “trust”-based decision processes had median response times for the consent question of 48 and 59 seconds respectively (see Table 2, Panel A). Particularly the combination of “consequentialist” and “trust” was associated with longer response times and presumably more thorough processing. In contrast, those respondents who reported “gut feeling” or “usual” decision processes had much shorter response times, with medians of 24 and 28 seconds respectively. Qualitatively similar results for response time were obtained in all samples.

A similar pattern can be observed in the respondent self-reports of the effort required for the decision (see Table 2, Panel B). Effort was reported using an 11-point scale ranging from 0 to 10, with higher values indicating higher effort. These data are only available in AP1.2 and AP2. As with response times, 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 “usual”.

Our third marker of effort is whether respondents consulted the additional materials on the data linkage process while deciding whether to consent. In the online surveys, this required clicking hyperlinks to access the leaflet or diagram. In the face-to-face survey (IP11A), the interviewer provided these materials and noted whether the respondent looked at them or not. These results are presented in Panel C of Table 2 and show similar patterns to the other markers of effort in Table 2. Respondents who based their consent decision on consequentialist thinking or trust made more use of additional materials than respondents who reported gut feeling or their usual behavior as their decision process. This is even more so if respondents used a combination of both, consequentialist and trust-based decision processes. Although the absolute levels of consultation vary quite substantially between the samples, the relative pattern across self-reported decision processes is the same for all samples.

Table 2: Markers of Decision Effort by Reported Decision Process

	IP11A	IP11B	AP1.1	AP1.2	AP2
	Panel A: Response Time (Median in Seconds)				
consequentialist	93	36	48	33	36
trust	-	-	59	42	-
consequentialist + trust	-	-	81	68	-
gut feeling	85	26	24	23	24
usual	76	29	28	26	30
Kruskal-Wallis rank sum test	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
(Number of groups)	(3)	(3)	(5)	(5)	(3)
	Panel B: Reported Effort, 0-10 scale (Means)				
consequentialist				5.9	6.2
trust				5.5	-
consequentialist + trust				5.8	-
gut feeling				4.7	4.9
usual				4.5	4.7
overall F-test				p<0.001	p<0.001
	Panel C: Consulted Leaflet or Diagram				
consequentialist	54%	15%	42%	39%	34%
trust	-	-	54%	48%	-
consequentialist + trust	-	-	72%	70%	-
gut feeling	44%	10%	21%	25%	20%
usual	38%	7%	20%	23%	16%
overall chi2 test	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001

Notes:

1. This table is based on an exclusive coding of the 5 most common patterns of decision processes. "Other patterns" are not presented.
2. In all web surveys, we assume that respondents consulted the additional material when they clicked on it. For IP11A, we assume that the respondent consulted the material if the interviewer reported to have "talked through all of the flowchart" or reported that the respondent "appeared to read all of leaflet".

Finally, recall that as a further check in the AP2 sample we also asked people what role alternative factors played in making their decision. Each item was rated on a five-point scale with 1 being *"played no role"* and 5 being *"played a very big role"*. Items such as *"what information the government has about me"*, *"how much I know about the organisations involved"*, and *"the benefits to society"* played a larger role for those respondents who reported a consequentialist decision process.

Table 3 shows the results by reported decision process. As the trust option was not asked in AP2, we focus on consequential, usual and gut responses.

The answers to this question also show that respondents who reported a consequentialist decision took a broader range of factors into account in making their

decision. We treat a factor as ‘considered’ if the respondent rated it as 4 or 5 on the scale. Respondents who reported a consequentialist decision considered on average 6.4 factors, while respondents who reported a gut- or habit-based decision considered on average 5.8 and 6.1 factors respectively (an overall F-test yields $F(2)=2.59$ and $p=0.076$). These results suggest that the decision process varied both in which type of information respondents based their decision on, and in the amount of information they considered.

Table 3: Decision Processes and Information Considered for the Consent Decision (AP2)

	Decision Process			
	consequentialist	**gut feeling**	**usual**	**p-value (overall F-test)**
Mean ratings (5 point scale, with 5 most important).				
INFORMATION				
What information the government has about me	3.6	3.2	3.1	p<0.001
How much information I've already shared	3.3	3.0	3.2	p=0.018
How much I know about the organisations involved	3.7	3.5	3.3	p=0.003
Who might have access to my information	4.0	3.7	4.0	p=0.006
RISKS				
How much I trust the organisations involved	3.9	3.7	3.6	p<0.001
Whether the Government would use this data to check on me	3.3	3.2	3.1	p=0.271
Recent news stories related to data security	2.7	2.7	3.1	p=0.002
The chances of my personal data getting into the wrong hands or being misused	3.8	3.6	4.0	p=0.002
BENEFITS				
Wanting to be helpful	3.2	3.0	2.6	p<0.001
The benefits to society	2.9	2.6	2.3	p<0.001
PROCESSES				
The way I usually make decisions	3.4	3.4	3.8	p<0.001
My gut feelings	3.2	4.2	3.7	p<0.001
What friends and family would do	2.0	2.1	2.0	p=0.328
Avoiding difficult decisions	2.2	2.6	2.4	p=0.008
Mean # of Factors Considered	6.4	5.8	6.1	p=0.076

Notes:

1. In this table we use the exclusive coding for the decision processes. As the trust option was not asked in AP2, we focus on the three of our five most common patterns that are relevant to AP2: consequential (alone), usual (alone) and gut (alone). F-tests stem from regressions of variables in first column (one per row) on the three decision processes. Respondents who reported "other" or less frequent patterns are omitted from the analysis.
2. In counting factors considered, we treat a factor as 'considered' if the respondent rated it as 4 or 5 on the five-point response scale.

In summary, these findings give us confidence that self-reported decision processes measure genuine differences in how respondents arrived at their answers to the consent question. Relative to those who report less-reflective processes (instinct or gut feeling, habit or usual practice), those who report using more-reflective processes (consequentialist and/or trust) took longer to answer the consent question, reported investing more effort in the decision, consulted the additional materials at a higher rate, and these patterns replicate across all samples. In addition, the nature and quantity of reported factors that play a role in their decision are consistent with the decision process respondents reported.

4.2 Determinants of the Decision Process

We next consider the extent to which the decision process is predicted by socio-demographic background characteristics (sex, age, education and employment status), and by survey design. As described above, we experimentally manipulated the difficulty of the question wording (for the AP1 and IP samples), the location of the consent question in the questionnaire (for the IP11A sample), and whether the respondents were shown a trust-priming statement (for AP2).

Table 4 reports average marginal effects estimated from logit models of the probability that the consent decision was processed in a consequentialist way. The dependent variable in these models is coded as one if respondents selected "consequentialist" as a response option, either alone or in conjunction with other options. Table 5, in turn, reports average marginal effects estimated from logit models of the probability that the self-reported decision process involved trust. The dependent variable in these models is coded as one if respondents selected "trust" as a response option, either alone or in conjunction with other options.

Overall, Tables 4 and 5 indicate that the decision process is not strongly predicted by sociodemographic variables or the survey design elements that we experimentally manipulated. All models have limited explanatory power with McFadden Pseudo-R^2 values ranging between 0.01 and 0.025.

The F-tests reported in Table 4 indicate that higher levels of education predict more consequentialist processing in all samples. Those with a university degree are seven to fifteen percentage points more likely to report a consequentialist process than the lowest education group. Effects for A-levels (an academic-track high school qualification) are of similar magnitudes, albeit less consistently statistically significant

at conventional levels. Table 5 indicates that older respondents (aged sixty and above) are more likely to report a trust-based decision process, by seventeen to twenty-two percentage points relative to those aged forty and below, and these effects are statistically significant ($p<0.001$).

Turning to survey design, easier wording does not induce more consequentialist or trust-based processing in any sample. Both the wording difficulty and question placement experiments were implemented in IP11A and the logistic regression model in column 1 of Table 4 allows for both main effects of these experiments and an interaction. We found no significant interaction of wording and placement on the decision process used.

Table 4: Logistic Regressions for "Consequentialist" Decision Process (reported alone or among others)

Average Marginal Effects (AMEs)

	(1) (IP11A)	(2) (IP11B)	(3) (AP1.1)	(4) (AP1.2)	(5) (AP2)
Background Characteristics					
sex (female)	-0.03	-0.06*	-0.03	-0.05	-0.07*
	(-0.93)	(-2.06)	(-0.95)	(-1.50)	(-2.31)
age (41-59)	-0.01	0.03	0.01	-0.02	-0.01
	(-0.36)	(0.94)	(0.28)	(-0.42)	(-0.18)
age (60+)	0.01	0.02	0.01	-0.04	0.01
	(0.40)	(0.50)	(0.36)	(-0.74)	(0.14)
Partial F-test	*p=0.77*	*p=0.64*	*p=0.93*	*p=0.76*	*p=0.95*
education (A(S)-level)	0.15***	0.09*	0.05	0.05	0.14**
	(3.41)	(2.21)	(1.16)	(1.02)	(3.24)
education (degree)	0.10***	0.07*	0.15***	0.14***	0.12***
	(3.48)	(2.31)	(4.24)	(3.73)	(3.39)
Partial F-test	*p<0.001*	*p=0.02*	*p<0.001*	*p<0.001*	*p<0.001*
in work	0.10**	-0.01	0.00	0.02	-0.03
	(3.10)	(-0.29)	(0.11)	(0.45)	(-0.87)
Design experiments:					
wording (easy)	0.01	0.05	-0.01	-0.01	
	(0.31)	(1.93)	(-0.25)	(-0.24)	
location (early)	0.04				
	(1.44)				
trust priming					-0.03
					(-0.84)
Additional Marginal Effects:					
wording if late location	0.07				
	(1.73)				
wording if early location	-0.05				
	(-1.20)				
N	1301	1227	1030	815	965
pseudo R2	0.024	0.010	0.015	0.021	0.016

Notes:

1. t statistics in parentheses
2. p < 0.05, ** p < 0.01, *** p < 0.001
3. The regression uses an inclusive coding of the decision process, meaning that "consequentialist" is coded as one whenever it is reported alone or in conjunction with any other response option, and zero otherwise. We exclude 64 IP respondents who did not select any response option, not even "other".

Table 5: Logistic Regressions for "Trust" Decision Process
(reported alone or among others)
Average Marginal Effects (AMEs)

	(1) AP1.1	(2) AP1.2
Background Characteristics:		
sex (female)	-0.08**	-0.04
	(-2.59)	(-1.22)
age (41-59)	0.05	0.11*
	(1.25)	(2.57)
age (60+)	0.17***	0.22***
	(4.12)	(4.57)
Partial F-test	*p<0.001*	*p<0.001*
education (A(S)-level)	0.04	0.04
	(0.95)	(0.84)
education (degree)	0.05	0.09*
	(1.44)	(2.25)
Partial F-test	*p=0.33*	*p=0.08*
in work	-0.03	0.04
	(-0.89)	(0.98)
Design experiments:		
wording (easy)	0.03	0.02
	(1.11)	(0.49)
N	1030	815
pseudo R^2	0.025	0.023

Notes:

5. t statistics in parentheses
6. p < 0.05, ** p < 0.01, *** p < 0.001
7. The regression uses an inclusive coding of the decision process, meaning that "trust" is coded as one whenever it is reported alone or in conjunction with any other response option, and zero otherwise.

4.3 Associations between Decision Process and Consent Outcomes

Table 6 reports how our four consent outcomes vary by self-reported decision process. Again, those four outcomes are the consent decision itself, comprehension of the request (measured objectively and subjectively) and confidence in the consent decision. We again contrast the five most common response options for decision process. Panel A shows that respondents who reported more reflective processes ("consequentialist" and/or "trust") had higher consent rates than those who reported less reflective processes ("gut feeling" or "usual"). This replicates across all five samples and 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 "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 "usual" have very low consent rates. For example, in AP1.1, 83.2% of those who report "trust" alone and 77.1% of those who reported "trust" together with "consequentialist" consented to record linkage, compared with only 19.8% of those who reported doing what they usually do.

Objective knowledge of the consent request (Panel B) follows a similar pattern. Recall that this measure 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 higher levels of knowledge than those who reported less reflective processes ("gut feeling" or "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. The same pattern is seen for subjective knowledge (Panel C). Respondents who reported "consequentialist"- or "trust"-based decision have higher mean self-rated understanding of between 0.2 and 0.6 points on a four-point scale, and again the differences are significant at conventional significance levels. Together these results suggest that those employing consequentialist and trust-based decision processes have more *informed* consent.

Interestingly, reported confidence in the consent decision is less correlated with the decision process than comprehension. Respondents who report basing their decision on what they "usually do" report high levels of confidence in that decision. The same respondents also show the lowest consent rates by far. Thus, this is a group of

respondents that confidently refuses consent and is likely difficult to persuade otherwise.

Table 6: Consent Outcomes by Reported Decision Process

	IP11A	IP11B	AP1.1	AP1.2	AP2
	Panel A: Consent Rate (Percentage)				
consequentialist	85.2%	66.5%	54.6%	52.9%	70.5%
trust	-	-	83.2%	88.7%	-
consequentialist + trust	-	-	77.1%	82.8%	-
gut feeling	75.0%	48.5%	42.9%	40.8%	54.9%
usual	52.7%	19.6%	19.5%	19.8%	28.6%
overall chi2 test	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
	Panel B: Objective Knowledge (Means)				
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	4.3
usual	4.2	3.8	4.1	4.1	4.3
overall F-test	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
	Panel C: Subjective Knowledge (Means)				
consequentialist	3.1	2.7	2.5	2.5	2.6
trust	-	-	2.6	2.6	-
consequentialist + trust	-	-	2.7	2.8	-
gut feeling	2.6	2.2	2.0	2.0	2.0
usual	2.9	2.3	1.9	2.1	2.1
overall F-test	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
	Panel D: Reported Confidence (Means)				
consequentialist	3.1	3.0	2.9	2.9	2.9
trust	-	-	2.7	2.8	-
consequentialist + trust	-	-	2.9	2.9	-
gut feeling	2.8	2.8	2.9	3.0	2.8
usual	3.0	3.1	3.1	3.1	3.1
overall F-test	p<0.001	p<0.001	p<0.001	p=0.04	p<0.001

Notes:

1. The objective knowledge score obtained from the sum of correct answers to 8 knowledge questions
2. Subjective knowledge is reported on a 4-point scale ranging from 1 to 4, with higher values indicating higher knowledge.
3. Confidence is reported on a 4-point scale ranging from 1 to 4, with higher values indicating more confidence.

Those surveyed in both AP1.1 and AP1.2 form a panel of respondents observed at two points in time a year apart (May 2018 and May 2019). This also allows us to

examine how stable decision processes are across consent decisions for a single individual, and to test for a longitudinal association between decision processes and consent outcomes. This is a natural first step in thinking about whether the associations documented in Table 6 are causal. A "fixed-effects" or "within-individual" estimator can eliminate time-invariant unobservables as potential confounders.

Between AP1.1 and AP1.2, 75% of respondents made some change in their reported decision process, and 24% of respondents changed their response to the request for consent to data linkage (13% changed from "yes" to "no" and 11% changed from "no" to "yes"). In Table 7 we document these transitions in more detail. Here for ease of exposition we group responses about the decision process into three categories: those who only report one or more reflective decision processes ("consequentialist" or "trust"-based, or both); those who only report one or more instinctive decision processes ("gut," "usual," or both); and finally a residual category that includes those who responded with "other", alone or in conjunction with other responses, and the small number of respondents who reported some mix of more and less reflective decision processes (for example, "trust" and "usual"). The rows of Table 7 correspond to decision process responses in AP1.1, while the columns correspond to responses in AP1.2.

Table 7: Stability of Decision Process and Consent

	AP1.2 (2019)		
	Consequentialist and/or Trust	Usual or Gut	Other patterns
Panel A: Stability of decision process (Row percentages)			
AP1.1 (2018):			
Consequentialist and/or Trust	56.0	24.1	19.9
Usual or Gut	28.9	49.5	21.7
Other patterns	24.2	27.1	48.8
Panel B: Consent unchanged (Cell percentages)			
Consequentialist and/or Trust	80.1	60.0	66.7
Usual or Gut	56.3	83.9	86.7
Other patterns	70. 0	85.7	84.2

Notes:

1. We use the exclusive coding that focuses on the five most common response patterns, but to maintain cell size when examining transitions, we collapse those groups into two. We combine trust alone, consequential alone and trust plus consequences into one group of reflective decision makers, and collapse usual (alone) and gut (alone) into a group of less reflective decision makers. All remaining cases are grouped into "other patterns".

Panel A of Table 7 summarizes the stability of decision processes. The number in each cell is the row percentage. So, for example, of those who said "consequentialist" or "trust" (or both) in AP1.1, 56.0% reported a more reflective decision process in AP1.2 as well, while 24.1% switched to less reflective or more instinctive decision process ("gut" or "usual") and 19.9% moved to the other patterns. Panel A indicates that decision processes are persistent, but not unchanging. Respondents who reported a reflective decision process in 2018 were more likely than not to do the same in 2019, and similarly those who reported a more instinctive decision process in 2018 were more likely than not to do the same in 2019. But there is significant mobility, with almost a quarter (24.1%) of respondents who reported a reflective decision process in 2018 reporting an instinctive decision process in 2019, and more than a quarter (28.9%) of

those who reported an instinctive decision process in 2018 reporting a reflective decision process one year later.

Panel B of Table 7 then considers the stability of consent decisions *within* each of these patterns of decision processes. In this panel, each number is a cell percentage. For example, the top left corner of Panel B indicates that of those who reported a reflective decision process in both 2018 (AP1.1) and 2019 (AP1.2), 80.1% gave the same consent decision (and hence 19.9% changed their consent decision, from "yes" to "no", or "no" to "yes'). Similarly, of those who reported an instinctive decision process in both 2018 and 2019, 83.9% gave the same consent decision. Those who switched from a reflective decision process to an instinctive process, or from an instinctive process to a reflective process had much less stable consent decisions (60.0% and 56.3% respectively). This panel indicates that changes in decision process are associated with changes in decision.

Table 8 reports estimates of linear "fixed-effects" models which document the longitudinal (or within-person) association between decision process and consent outcomes (columns 2 through 4), as well as between decision process and the logarithm of response time (column 1).[5] These estimates use a "within" transformation to eliminate additive, time-invariant unobservables that might confound associations estimated in the cross-sectional data. The independent variables are two binary indicators: one for whether the respondent reported a consequentialist decision and one for whether they reported a trust-based process. The two indicators are each coded as 1 if mentioned and 0 otherwise and are not mutually exclusive.

The first result in Table 8 is that, within individuals over time, more reflective decision processes (consequentialist and trust-based) are associated with longer response times. This further strengthens the case that self-reported decision processes are capturing real differences in decision making. Turning to consent outcomes, Table 8 indicates that the strong association of trust-based processing with consent, previously documented in the cross-sectional data, is also found longitudinally, within respondents (and even with this relatively modest sample size). The same is not true however of the association between consequentialist processing and consent. Subjective knowledge is

[5] The consent decision is a binary outcome, objective and subjective knowledge are, respectively, 9- and 4-point scales, and log response time is a continuous variable. Nevertheless, we report a linear model in all cases. Using a conditional logistic regression (allowing for individual fixed effects) for the binary consent variable does not change our conclusions.

strongly associated with both trust-based and consequentialist processing within respondents over time. Longitudinal associations with objective knowledge are weaker, though for trust-based processing the effect size is substantively meaningful, and marginally statistically significant.

Table 8: Linear Fixed Effects Models of the Effect of Decision Mode on Consent Outcomes and Response Time.

	(1) log response time (xtreg)	(2) consent (xtreg)	(3) obj knowledge (xtreg)	(4) subj knowledge (xtreg)
consequentialist	0.217**	-0.00123	0.0628	0.193**
	(2.89)	(-0.04)	(0.61)	(3.07)
trust	0.173*	0.225***	0.210*	0.346***
	(2.49)	(8.46)	(2.21)	(5.98)
_cons	3.500***	0.391***	4.403***	2.102***
	(77.12)	(22.44)	(70.77)	(55.49)
N	1632	1632	1632	1632
N (individuals)	816	816	816	816
R^2	0.017	0.081	0.006	0.052

Notes:
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Discussion and Conclusions

Our analyses indicate that survey respondents report using different decision processes, some of which are more "reflective" (for example, considering the consequences of consent or their trust in the organisations involved) while others suggest the use of less or different information (for example, based on "gut-feeling"). Across all samples, fewer than 40% of our respondents report using a reflective decision process exclusively. These self-reported processes are corroborated by markers of effort: those who report a more reflective processing of the request take longer to respond, are more likely to read additional information about the linkage and are more likely to self-report higher levels of effort in answering the question. Reported decision processes align with the nature and quantity of information respondents report drawing on in making the decision. Self-reported process is weakly predicted by background characteristics and not affected by our survey design manipulations.

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. It is notable that we find these associations consistently across different samples and interview modes. Respondents in the IP samples (which are part of a longitudinal study) have an existing relationship with the survey organisation and have been asked for consent to data linkage before, and so bring a history to the request. In contrast, AP participants are not used to requests like this. Despite this, the samples give very similar results.

Many attempts to achieve higher consent rates and more informed consent involve the provision of additional information, and as our review of the literature indicates, these have had limited success. Our analyses point to an explanation: if many of those who withhold consent are using a very rapid and less reflective 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 organisations. Our results indicate that if it were possible to shift respondents to more reflective decision processes, this could have a double benefit: in our data those employing the more reflective decision processes are *both* more likely to consent and have greater understanding of the request. That is, they are more likely to give informed consent.

Our analysis has several limitations, which in turn suggest important directions for future research. First, while we found that respondents were able to answer a question about their decision processes, our decision process question was a first attempt to elicit such information, and a useful next step would be to refine this kind of questioning. Second, we do not claim a causal interpretation for the associations we find between reflective decision processes and consent, and reflective decision processes and understanding. The longitudinal associations we document in the AP eliminate selection on time invariant unobservables as an alternative explanation, but we remain cautious. Alternative designs to establish causality should be a priority for future research. Related to this, more work is needed to investigate if and how more reflective decisions processes can be encouraged. Our experimental manipulations were not effective in altering respondents' choice of decision process. "High hurdle" techniques are sometimes used in the medical world to push individuals to more

reflective decision making by, for example, requiring the individual to satisfy a comprehension test before giving consent, but it is not clear how such techniques could be adopted for time-constrained and low-risk contexts such as surveys. Further work on understanding how respondents select a decision process and how that choice can be influenced is needed, but such work is likely to involve smaller scale lab studies at least initially.

Funding and Declarations of Interest.

This research was funded by the Nuffield Foundation (www.nuffieldfoundation.org) with co-funding from the Economic and Social Research Council (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 views expressed here are those of the authors and not necessarily of the Nuffield Foundation or the ESRC.

Declarations of interest: None.

References

American Association for Public Opinion Research (AAPOR) (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 9th edition. Washington, DC: AAPOR.

Beninger, K., Digby, A., and MacGregor, J. (2017), "Understanding Society: 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 Series 2020-02,* Colchester: University of Essex.

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.

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.

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 Tang, S. (2017), *Longitudinal Studies Strategic Review: 2017 Report to the Economic and Social Research Council* Swindon: ESRC.

Desch, K., Li, J., Kim, S., Laventhal, N., Metzger, K., Siemieniak, D., and Ginsburg, D. (2011), "Analysis of informed consent document utilization in a minimal-risk genetic study." *Annuals of Internal Medicine*, 155: 316–322, https://doi.org/10.7326/0003-4819-155-5-201109060-00009

Edwards, B., and Biddle, N. (2021), "Consent to data linkage: Experimental evidence from an online panel." Chapter 8 in P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*. New York: Wiley.

Galotti, K.M. (2007), "Decision structuring in important real-life choices." *Psychological Science*, 18 (4): 320-325.

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, https://doi.org/10.1525/jer.2013.8.3.12

Gigerenzer, G. (2000), *Adaptive Thinking: Rationality in the Real World*. New York: Oxford University Press.

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.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011), "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review*, Vol. 105(4):765-789.

Jäckle, A., Al Baghal, T., Burton, J., Kaminska, O., and Lynn, P. (2019), *Understanding Society: The UK Household Longitudinal Study Innovation Panel, Waves 1-11, User Manual*. Colchester: University of Essex.

Jäckle, A., Beninger, K., Burton, J., and Couper, M.P. (2021a), "Understanding data linkage consent in longitudinal surveys." Chapter 6 in P. Lynn (ed.) *Advances in Longitudinal Survey Metho*dology. New York: Wiley.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F. and Walzenbach, S. (2021b), "Understanding and improving data linkage consent in surveys." *Understanding Society Working Paper Series 2021-01*, Colchester: University of Essex.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., and Walzenbach, S. (2021c). "How and why does the mode of data collection affect consent to data linkage?" *Understanding Society Working Paper 2021-04*, Colchester: University of Essex.

Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F., and Walzenbach, S. (2021d). "Consent to data linkage: Question wording and format experiments." *Draft*. Colchester: University of Essex.

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*. Vol. 169(4): 701-722.

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

Kahneman, D., Slovic, P., and Tversky, A. (eds.) (1982), *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Keele, L., Tingley, D. and Yamamoto, T. (forthcoming), "identifying mechanisms behind policy interventions via causal mediation analysis." *Journal of Policy Analysis and Management*.

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.

Krosnick, J.A. (1991), "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology*, 5: 213-236.

Krosnick, J.A., and Alwin, D.F. (1987), "An evaluation of a cognitive theory of response-order effects in survey measurement." *Public Opinion Quarterly*, 51(2): 201-219.

McNutt, L.-A., Waltermaurer, E., Bednarczyk, R.A.,Carlson, B.E., Kotval, J., McCauley, J., Campbell, J.C., and Ford, D.E. (2007), "Are we misjudging how well informed consent forms are read?" *Journal of Empirical Research on Human Research Ethics*, 3(1): 89-97, https://doi.org/10.1525/jer.2008.3.1.89

Meyer, B.D., Wallace K.C.M., and Sullivan, J.X. (2015), "Household surveys in crisis." *Journal of Economic Perspectives*, 29(4):199-226.

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.

Peycheva, D., Ploubidis, G., and Calderwood, L. (2021), "Determinants of consent to administrative records linkage in longitudinal surveys: Evidence from Next Steps." Chapter 7 in P. Lynn (ed.) *Advances in Longitudinal Survey Methodology*, Chichester: Wiley.

Petty, R.E., and Cacioppo, J.T. (1986), "The elaboration likelihood model of persuasion." *Advances in Experimental Social Psychology*, 19: 123-205.

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., Schmucker, A., Kreuter, F., Couper, M.P., and Holtmann, L. (2021), "Respondent understanding of data linkage consent." *Survey Methods: Insights from the Field*, retrieved from https://surveyinsights.org/?p=15407.

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.

Simon, H.A. (1957), *Models of Man: Social and Rational*. New York: Wiley.

Simon, H.A. (1959), "Theories of decision making in economics." *American Economic Review*, 49: 223-283.

Strack, F., and Deutsch, R. (2004), "Reflective and impulsive determinants of social behavior." *Personality and Social Psychology Review*, 8: 220-247.

Tversky, A., and Kahneman, D., (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185: 1124-1130.

University of Essex, Institute for Social and Economic Research (2021), Understanding Society: Innovation Panel, Waves 1-13, 2008-2020. [data collection]. 10th Edition. UK Data Service. SN: 6849, http://doi.org/10.5255/UKDA-SN-6849-14.

Walzenbach, S., Burton, J., Couper, M.P., Crossley, T.F., and Jäckle, A. (2021), "Experiments on multiple requests for consent to data linkage in surveys." *Understanding Society Working Paper Series 2021-02*, Colchester: University of Essex.

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." Presented at the 4th Panel Survey Methods Workshop, Ann Arbor, MI.

Online Appendix A: Survey Response Rates

Innovation Panel Wave 11

The Wave 11 household response rate for the continuing samples was 73.2%, with 80.5% of individuals in responding households giving full interviews. For documentation of the IP survey design and implementation see Jäckle et al. (2019). The achieved sample size for IP11 was 2,896 respondents, of which 2,662 were used in analyses (after dropping a few cases who had difficulty with the CASI module).

IP11A (face-to-face): 80.5% of households allocated to FTF-first (and 80.8% of individuals in those households) responded.

IP11B (web): 77.6% of households allocated to web-first (and 83.2% of individuals in those households) responded.

Access Panel

AP1.2: A total of 46,206 panellists were invited to the first survey, of whom 6,532 started the survey and 5,633 completed it (401 broke off and 498 were screened out), for a participation rate of 12.2% (see AAPOR 2016).

AP1.2: Among those who completed AP1.1 and were invited to the second wave (AP1.2), a total of 2,064 respondents completed AP1.2 for a conditional response rate of 36.6%.

AP2: A total of 30,682 panellists were invited to the survey, of whom 6,459 started the survey and 3,850 completed it (301 broke off and 2,308 were screened out), for a participation rate of 12.5%.

The AP samples included a number of different experiments. We use the data from a subset of cases asked a single consent question.

Online Appendix B: Question Wording

B1. Standard (control) consent question

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 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

B2. Easy consent question

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

B3. Trust prime

(The statement below preceded the easy wording consent question for both the prime and no prime (control condition.)

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

B4. 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]

(Correct responses [indicated in brackets] summed to create index of objective knowledge, ranging from 0-8.)

B5. 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

B6. 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

B7. 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