Understanding Society

Working Paper Series

No. 2016 – 06

August 2016

Is that still the same? Has that changed? On the accuracy of measuring change with dependent interviewing

Annette Jäckle

Institute for Social and Economic Research, University of Essex

Stephanie Eckman

RTI International

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

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

Non-technical summary

One of the main purposes of panel surveys is to measure the changes, and stability, experienced by sample members. Changes however tend to be over-reported due to various forms of measurement error and therefore panel surveys often remind respondents of their answer in the previous interview before asking about their current situation (referred to as proactive dependent interviewing). There is however concern that respondents may minimize the effort of answering survey questions by simply confirming previous information as true, which could lead to under-reporting of change. In this study we examine whether it matters how proactive dependent interviewing questions are worded, whether respondents exploit possible least-effort strategies, and which wording provides the most accurate data on changes.

We experimentally test different wordings where the respondent is reminded of their previous answer and then for example asked *"Is this still the case?,"* or *"Has this changed?,"* or *"Is this still the case or has it changed?."* We also compare with independent questions where respondents are not reminded of their previous answer. We find that all question wordings reduce the reporting of change compared to independent interviewing. We do not find evidence that respondents make use of the reminder for least effort strategies: we find no evidence suggesting that they simply answer "yes" to yes/no questions regardless of the question content, or that they deliberately report no change to avoid follow-up questions. We also find no evidence that the question wording primes respondents to either think about things related to the question that have changed or not changed, depending on the wording. We also link the experimental data to administrative records of our sample members, to examine which wording provides most accurate data. Overall our findings suggest that asking *"Is that still the case?"* is the best wording for proactive dependent interviewing questions.

Is that still the same? Has that changed? On the accuracy of measuring change with dependent interviewing

Annette Jäckle (University of Essex)

Stephanie Eckman (RTI International)

Abstract:
We experimentally test how best to word proactive dependent interviewing questions, whereby respondents are reminded of their answer in the previous interview, before being asked about their current situation. First we test for differences in reporting of change between possible wording formats including independent interviewing, second we examine potential mechanisms through which the wording may affect responses, and third we link to administrative records to assert which wording provides the most accurate measures of change. The overall results suggest that reminding respondents of their previous answer and then asking *"Is that still the case?"* produces the most accurate data on change and stability experienced by individuals.

Keywords: panel survey, panel data, measurement error, record linkage, validation, proactive dependent interviewing

JEL classification: C83

Acknowledgements: The Measuring Change survey was funded by the Institute for Employment Research (IAB), Germany and most of the work on the survey and analysis was completed while Eckman worked at IAB. Jäckle gratefully acknowledges funding from IAB and from the Economic and Social Research Council for *Understanding Society* and the Innovation Panel (grant numbers RES-586-47-0001 and RES-586-47-0002). We thank Antje Kirchner for excellent support in implementing the Measuring Change study and Frauke Kreuter, Ashley Amaya, Ruben Bach, Theresa Goecke, Sarah Gorenflos, Charles Lau and Max Reinwald for comments on earlier drafts.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK, aejack@essex.ac.uk.

1 Introduction

The aim of panel surveys is to measure change: respondents are asked the same core set of questions at regular intervals, and changes in answers over time are interpreted as changes in respondents' situations. However, unless measurement error is perfectly correlated over time, the estimated changes reflect not only true change, but also changes in errors; as a result, panel surveys often over-estimate change. The method most commonly used to reduce spurious changes in panel data is dependent interviewing, whereby answers from the previous interview are preloaded and used in later questionnaires. Dependent interviewing is employed in most panel surveys using computer assisted interviewing to collect factual information such as labor market status, employment characteristics or income sources (e.g. in the UK Household Longitudinal Study: *Understanding Society*, the Current Population Survey, the National Longitudinal Survey of Youth 1997, the Health and Retirement Study, the English Longitudinal Study of Ageing, and the German panel study Labor Market and Social Security). The aim of this article is to examine how best to word dependent interviewing questions to maximize the accuracy of measures of change.

With proactive dependent interviewing, the respondent is reminded of his or her answer from the previous wave. For example: *"Last time we interviewed you on <date of interview>, you said you were <an employee/self-employed>."* This reminder is followed by a question about the current status, which can be worded in different ways, for example *"Is that still the case?"* or *"Has that changed?"* Alternatively, the survey could remind the respondent and then simply ask the original question again: *"Last time we interviewed you on <date of interview>, you said you were <an employee/self-employed>. What about now -- are you an employee or self-employed?"* (see Jäckle 2009). Proactive dependent interviewing is also used to reduce

redundancies and improve the flow of the interview: respondents whose situation has not changed can be routed around sections of the questionnaire.[1] In this paper, we focus on the effects that the wording of proactive dependent interviewing questions has on measurement error in reports of change.

Previous studies have shown that proactive dependent interviewing is effective in reducing spurious transitions in welfare program participation (Lugtig and Jäckle 2011; Lynn et al. 2012), ownership of assets and liabilities (Hoogendoorn 2004), and employment characteristics such as industry and occupation, managerial duties, or the size of the employing organization (Hill 1994; Lynn and Sala 2006; Perales 2014). Proactive dependent interviewing also reduces seam effects, the heaping of transitions in the interview month between recall periods in panel surveys, in labor market histories (Jäckle and Lynn 2007; Lemaitre 1992; Murray et al. 1991) and welfare receipt histories (Moore et al. 2009), and improves estimates of spell durations (Hill 1994; Jäckle 2008).

The benefits of proactive dependent interviewing come at a potential cost however. Some researchers and practitioners worry that reminding respondents of previous answers and asking whether this is still the case invites satisficing: "yes" is an easy and credible answer and so respondents may falsely confirm the previous information as still applying (see discussions by Holmberg 2004; Hoogendoorn 2004; Lynn et al. 2012; Perales 2014). If so, the original problem of spurious change would be replaced by spurious stability, introducing a new type of error in the estimation of change. This concern is not unfounded. There is indeed evidence that respondents

[1] Dependent interviewing can also be implemented reactively (Corti and Campanelli 1992). In this case, the respondent is not reminded of his or her previous answer but is asked the original question again at the later wave. The questionnaire script compares the two answers and, if a change is detected, prompts a follow-up question to verify the change. Reactive dependent interviewing is more commonly used for numeric questions such as income amounts, while proactive dependent interviewing is more commonly used for closed questions.

falsely confirm previous information: in two panel surveys where incorrect data were preloaded for proactive dependent interviewing questions (wave 9 of the National Longitudinal Survey of Youth (Aughinbaugh and Gardecki 2008) and wave 4 of the panel study Labor Market and Social Security (Eggs and Jäckle 2015)), non-random sub-samples of respondents were presented with an incorrect preload, which the majority did not correct. Experimental laboratory studies also provide evidence that dependent interviewing may indeed reduce spurious change by increasing spurious stability (Conrad, Rips and Fricker 2009; Rips, Conrad and Fricker 2003). What is unclear however is to what extent proactive dependent interviewing leads to satsisficing, whether the reduction in spurious changes outweighs any increase in spurious stability, and to what extent this trade-off is influenced by the wording of dependent interviewing questions.

How proactive dependent interviewing questions are worded varies between surveys – and often also within a survey. For example the UK Household Longitudinal Study uses the *"Has this changed?"* wording for some questions and *"Is this still the same?"* for others. The choice of question wording seems to be a matter of personal preference and taste. To date there has been little research into wording effects in proactive dependent interviewing questions (see Mathiowetz and McGonagle 2000). The exception is Al Baghal (2016) who concludes that asking about change is problematic: in face-to-face interviews respondents are more likely to say "yes" a change has occurred, but then frequently report the same status in response to follow up questions about their current status. However in the web version of the survey, the question wording does not affect reporting of change. This raises questions about the mechanisms through which the wording of dependent interviewing questions affects reporting of change, and how wording influences the accuracy of reports. We use experimental data and linked administrative records to examine the following questions:

(1) Does the wording of proactive dependent interviewing questions affect the likelihood of reporting a change?

(2) If yes, what are the mechanisms through which question wording affects the reporting of change?

(3) Which question wording produces the most accurate measures of change?

2 Background: How question wording may affect reporting of change

As many studies have shown, minor changes in how questions or response options are worded can influence respondents' answers (Krosnick and Presser 2010; Schwarz 1999), and thus we suspect that the different ways of asking proactive dependent interviewing questions can affect whether respondents report change or not. There are several potential mechanisms that could lead to a wording effect.

Satisficing leads respondents to pick an easy credible answer, instead of processing the question optimally and answering truthfully (Knowles and Condon 1999; Krosnick 1991). This mechanism may mean answering "yes" to all yes/no questions, or, on the phone, choosing the last answer choice heard when confronted with a long list (Holbrook et al. 2007). *Acquiescence*, the tendency to agree or to answer "yes" to yes/no questions, is a particular manifestation of satisficing which can be related to a respondent's personality and desire to please (Couch and Keniston 1960). This mechanism would also lead respondents to respond "yes" to all yes/no questions such as *"Is that still the case?"* and *"Has that changed?*". Alternative formats, which do not use a yes/no question, would not be affected by acquiescence.

Respondents may *learn* that they can reduce the number of questions by answering in certain ways, a phenomenon called *motivated misreporting*, which has been found in responses

to screener, filter and looping questions (Eckman and Kreuter 2015; Eckman et al. 2014; Kreuter et al. 2011; Tourangeau, Kreuter and Eckman 2012; Tourangeau, Kreuter and Eckman 2015). With proactive dependent interviewing, reporting that previous information still applies lets the respondent skip over questions about their current status. Respondents may falsely report no change to shorten the length of the questionnaire, leading to underreporting of change in all question wordings.

Biases in *decision making heuristics* may also affect respondents' answers. When deciding between alternatives, people tend to think of arguments in favor of the default before thinking of (fewer) arguments against it or in favor of the alternative (Dinner et al. 2011). As a result, choices between alternatives are biased towards the default. Similarly, asking *"Is this still the case?"* may prompt respondents to think about aspects related to the question that have not changed, leading them to conclude that their situation is still the same. Asking *"Has this changed?"* may prompt respondents to think about aspects that have changed, leading them to indicate a change.

Interviewer behavior may also lead respondents to falsely confirm preload information. Pascale and McGee (2008) examined audio-recordings of interviews from the English Longitudinal Study of Ageing and found that interviewers frequently deviated from the script, turning proactive questions into statements and thereby inviting confirmation.

Distinguishing the mechanisms that cause the dependent interviewing formats to perform differently is important. If motivated misreporting is the problem, we could design the questionnaire such that reporting no change does not reduce the number of questions (see, Hoogendoorn 2004). If the problem is related to acquiescence, we could avoid the use of one-

sided yes/no questions and instead use the balanced *"Is it still the same or has it changed?"* format, or follow the reminder with the original independent question.

In this paper we test whether satisficing, acquiescent response styles, motivated misreporting and biases in decision making heuristics affect the reporting of change with dependent interviewing questions. We examine the influence of interviewer behaviors in a companion paper (Jäckle, Al Baghal and Eckman 2015).

3 Data

We use data from three experimental studies to address our research questions: waves 3 and 7 of the UK Innovation Panel and a two-wave survey in Germany, which we refer to as the "Measuring Change" survey. Table 1 documents which study is used to address which research question. Below we describe the surveys and the experiments conducted in each.

The Innovation Panel survey

The Innovation Panel (IP) is part of the UK Household Longitudinal Study: *Understanding Society*, funded by the Economic and Social Research Council and led by the Institute for Social and Economic Research at the University of Essex. The IP is a stratified, clustered sample of Great Britain residents (for details on the sample design, see Lynn 2009). In this analysis, we use wave 3, fielded April to July 2010 by NatCen Social Research, and wave 7, fielded May to October 2014 by TNS BMRB. Wave 3 was conducted in person, and wave 7 used an experimental mixed-mode approach; however, we use only the cases assigned to the in-person mode. (We report on the implications of mixed modes for dependent interviewing in a separate paper, see Jäckle, Eckman and Nichols 2014.) For the cases fielded in wave 3, the household

response rate was 66.7%, with 82.2% of eligible individuals in these households responding. For cases fielded in-person in wave 7, the household response rate (for the wave 1 sample and the wave 4 refreshment) was 74.9%, with 81.5% of individuals in those households responding. (all response rates calculated using AAPOR's RR1, see The American Association for Public Opinion Research 2016).

IP3 Experimental design:

In wave 3 of the Innovation Panel (IP3), primary sampling units were randomly allocated to one of two groups, such that all adults in each PSU received the same treatment. Both groups were reminded of their answer in the previous interview, but the question to ascertain their current status varied: half received the STILL format and half the CHANGED format (see Table 2). Those who reported change were then asked follow-up questions about their current status. It was possible for respondents to report that a change had taken place, but then in the follow-up questions report that the same status still applied. We refer to the initial report of change, that is, the answer to the STILL or CHANGED question itself as *reported change*, and the subsequent report of change, that is, whether the status reported in wave 3 in fact differed from that reported in wave 2, as *corrected change*.

The experiment was implemented on four questions, each of which had been asked as an independent question in the previous wave: general health, whether job is permanent, and working hours (asked separately of employees and self-employed). See Appendix Table 1 for the full question wording. The data include observations on 2,577 dependent interviewing questions,

nested in 1,299 respondents.[2] The number of questions answered by each respondent varied due to routing.

IP7 Experimental design:

Wave 7 of the Innovation Panel (IP7) tested four versions of the dependent interviewing questions: the STILL format, the CHANGED format and two forced choice formats, as shown in Table 2. Households were randomly allocated to groups such that all respondents within a household received the same treatment. The experiment was implemented on 13 items in the household and individual questionnaires, which had each been asked as independent questions in the previous wave.[3] See Appendix Table 2 for question wordings. The in-person interviews were audio-recorded and we report on analyses of the recorded data in a separate paper (Jäckle, Al Baghal and Eckman 2015). The in-person data include observations on 1,903 dependent interviewing questions, nested in 474 respondents.[4]

The Measuring Change survey

The Measuring Change (MC) survey, a telephone study, was funded by the Institute for Employment Research (IAB) in Germany and fielded by the LINK Institute. Wave 1 was fielded September to November 2011 and wave 2 one year later. A national sample of adults was drawn from administrative records held by the German Federal Employment Agency (IAB 2011). The sample design has been described in more detail elsewhere (Eckman et al. 2015). The response

[2] After excluding 135 proxy respondents, 311 non-respondents in the previous wave, and 14 items with response discrepancies that could not be resolved.
[3] The experiment included a further six questions for which dependent interviewing had already been used in the previous wave. These items are excluded from our analyses.
[4] After dropping 45 proxy respondents, 166 previous wave non-respondents, 4 respondents who completed the survey by telephone, and 87 by web in the final non-response follow-up, and 14 items with response discrepancies.

rate at Wave 1 was 19.4% and the conditional response rate at Wave 2 was 63.2% (RR1, The American Association for Public Opinion Research 2016).

Respondents were asked for consent to link their survey data to the administrative records from which the sample was drawn. Of the 1,325 wave 2 respondents, 96% consented and were successfully linked. We exploit this link between survey and administrative data to answer the third research question about accuracy in reports on change. The dependent interviewing questions that could be validated with records were employment status (full-time, part-time, mini job[5], or other labor market activity) and receipt of two types of unemployment benefit (unemployment insurance, income support). The employment data stem from the German government's database of employer social security contributions (IAB 2013). All contributing jobs should be captured in the database; non-contributing positions, such as civil servant, police officer, professor, and the self-employed, are not covered (Jacobebbinghaus and Seth 2007). Due to these exclusions, we expected some mismatch between the responses and the data, but this error should be similar across the experimental groups. The administrative records on unemployment benefit contain information about all spells of unemployment benefit receipt. These data are of high quality as they are directly produced by the software that administers benefit claims and payments (Jacobebbinghaus and Seth 2007; Köhler and Thomsen 2009).

MC Experimental design:

The MC survey replicated both yes/no versions and one of the forced choice versions asked in IP7. In addition, the experiment included a version where respondents were reminded of their previous answer and then asked the independent question. A comparison with independent

[5] A mini-job is a form of German employment which was paid a maximum of 400 euros per month at the time of the survey and is not eligible for some employment benefits.

interviewing was also added. Respondents were randomly allocated to one of five treatment groups (see Table 2). The experiment was replicated in three modules: socio-demographics, labor market, and income sources. See the Appendix Table 3 for the question wording. The data include observations on 15,868 questions nested in 1,325 respondents.

4 Methods

The three data sets allow us to answer our three research questions, as specified in Table 1. Before proceeding with the analyses, we verified that the assignment of respondents to the various question wording conditions was in fact random: the cases do not differ in demographics, substantive variables, consent to record linkage or measures of change derived from the administrative data in any of the data sets.

All our analyses are performed at the item level rather than the respondent level. To account for the artificial inflation in the case base, we control for the clustering of items in respondents in all our models and significance testing using Taylor Series linearization.

For research question 1, about whether the wording of the dependent question affects reports of change, we present our main results in a figure, and use logistic models to test for significant differences between the formats. The dependent variable in these models is an indicator of change, coded as 1 if the answer changed between waves and 0 otherwise, and the independent variable is the question wording condition. We discuss the model results in the text, and the models themselves (estimated coefficients, fit statistics, case base) are given in Appendix Table 4.

For research question 2, we test four mechanisms – satisficing, acquiescence, motivated misreporting, and decision making heuristics – which offer competing explanations for how

question format may affect the reporting of change. The tests of satisficing, acquiescence and motivated misreporting make use of items in the IP3 and MC surveys that ask about personality traits and life situations and test memory and cognitive skills; these variables may moderate the mechanisms. Table 3 documents the relevant questions in IP3 and IMC and how we coded them. The distributions of the indicators are given in Figure 1 for IP3 and in Figure 2 for MC.

Satisficing would lead respondents to say "yes" in response to yes/no questions, because "yes" is a cognitively easier response to give (Knowles and Nathan 1997). This strategy would result in underreporting of change in the STILL format and overreporting in the CHANGED format. Respondents with low cognitive ability (measured by indicators 1, 2, 3, 4, 5 in Table 3), those less willing to expend cognitive effort (6, 7, 8), and those who were cognitively distracted during the interview (9, 13) should be more likely to satisfice (Krosnick 1991), and hence should have larger differences in reported change between the STILL and CHANGED formats, than respondents with high ability, high effort and those who were not distracted.

The *acquiescent response style* would also lead respondents to say "yes" more often than is warranted, which would again appear as underreporting of change in the STILL format and overreporting in the CHANGED format. With this mechanism, however, the moderator variables that we use to test the mechanism are different: we should see larger differences between question formats among respondents who have more agreeable personality traits (Couch and Keniston 1960) (10) or who acquiesce on other items in the survey (11,12).

We hypothesize that the variables shown in Table 3 are moderators that affect the relationship between the wording of the dependent interviewing question and the response given. We test the satisficing and acquiescence mechanisms by estimating moderator effects (Baron and Kenny 1986). Each test compares only the STILL and CHANGED formats, where it is clearest

what the effects should be. We estimate separate logistic models for each moderator variable, where the dependent variable is an indicator of reported change between waves and the independent variables are the moderator variable, the question wording format (STILL or CHANGED), and the interaction between the two. We interpret significant interactions in the expected direction as evidence of the hypothesized mechanism. We run separate models for each moderator variable and each dataset, IP3 and MC. Results from full models including all variables interacted with format are substantively and statistically similar to the results presented here.

Motivated misreporting predicts underreports of change with all wordings, as respondents seek to shorten the questionnaire: reporting no change lets respondents skip over the follow up questions about their current status. This strategy should be more common among respondents who are very busy (13) and thus likely to prefer a shorter interview, and among those who have high cognitive ability to think through the implications of their answers for later questions (1, 2, 3, 4, 5). However, because the mechanism involves lying, respondents who complete the interview in the presence of other members of their household (14) are probably less likely to underreport change to shorten the questionnaire. This expectation builds on previous evidence that respondents give more honest answers when others are present during the interview. For example, Aquilino (1993) finds that spousal presence produces more reports about pre-marital cohabitation and marital conflict; Silver, Anderson and Abramson (1986) show that the presence of children and spouses improve the quality of voting data. We test the motivated misreporting mechanism using the same logistic models, but interpret them differently. In this test, significant main effects on the explanatory variables in the expected direction along with non-significant

interactions are evidence of the hypothesized mechanism, because the strategy should work the same in the STILL and CHANGED formats.

We test the *decision heuristics* mechanism in a different way. If this mechanism plays a role, we expect respondents who are asked the STILL version to be more likely to think of aspects of their situation that have remained the same, and respondents who were asked the CHANGED version to think about changes. We therefore added open-ended follow-up questions to some of the dependent interviewing questions in the MC survey, asking respondents what had gone through their minds when thinking about how their current situation compared to their situation at the previous interview (moderator 15 in Table 3).[6] We use the coded open responses to derive two dependent variables: the first variable is coded as 1 if the respondent mentioned something that had stayed the same since the previous year, and 0 otherwise; the second variable is coded as 1 if the respondent mentioned something that had changed or was expected to change, and 0 otherwise. To test for differences in decision heuristics between question wordings, we estimate logistic models using these two dependent variables and the question format indicators as explanatory variables.

For research question 3, we exploit the link between the MC survey responses and administrative records, with which we can check the accuracy of reports of change in three items in the survey: receipt of unemployment benefits, receipt of income support, employment status. The first two are already binary indicators and we converted employment status into binary indictors as well: full-time employment, part-time employment and mini-job.[7] For each of these five variables, we coded whether the state applied to each respondent in each wave, according to

[6] See Appendix Table 5 for the coding frame that we used to code 1,902 open answers, nested in 1,200 respondents.
[7] A mini-job is a form of German employment which was paid a maximum of 400 euros per month at the time of the survey and is not eligible for some employment benefits.

the survey data and the administrative data. The end result is ten transition variables: one for each of the five variables in the two sources. To illustrate, "No-No" for full-time employment from the survey data indicates that the respondent reported no full-time job in both wave 1 and wave 2. "Yes-No" in the income support variable from the administrative data means that the records indicate the respondent was receiving income support at the time of wave 1 and was no longer receiving support at wave 2. These transition variables let us test how the accuracy of reports of change vary with the DI question wording.

5 Results

Does the wording of proactive dependent interviewing questions affect the likelihood of reporting a change?

Figure 3 shows the rates of change in the three studies by question wording. There are two types of change shown in the figure. *Reported change* refers to the answers to the five different dependent interviewing question formats: these change rates are shown with the dark bars in Figure 3. If respondents indicate a change, they are then asked the independent question about their status. *Corrected change* takes the answer to this follow-up into account. In some cases, respondents indicated that there had been a change, but then reported the same answer in the second wave as they had in the first: these are the light bars in the figure, and thus corrected change may be lower than reported change. Because the rates of change differ with the topic of a question and the population surveyed, we concentrate on comparisons within each study and on the patterns across the three studies, rather than on the corrected rates.

In IP3, the STILL and CHANGED formats clearly work differently, with CHANGED gathering many more reports of change: the difference between the formats is highly significant

(β=1.069, p=0.000; the logistic models used to test significance of differences between wordings are given in Appendix Table 4). The difference in corrected rates of change between question wordings (light grey bars in Figure 3) is smaller, but still significant (β=0.404, p=0.001).

There are several possible explanations for the differences between reported and corrected change. Respondents may be satisficing or acquiescing, giving the easy answer, "yes", in response to the *"Has that changed?"* question and then backtracking when they realize that was not the right answer. It is also possible that they are reporting correctly. Consider the question: *"In the last interview, you said that you were full-time employed. Has that changed?"* A respondent may think about the job she had last year and the fact that she lost that job and got a new one, which is also full-time. Then from the respondent's perspective, a change has taken place in the intervening year, but from the survey's perspective, the correct answer is that no change has taken place, because the respondent is employed full-time at the time of both interviews. The CHANGED format may be particularly susceptible to this error, because it primes the respondent to think about change. Interviewer errors are a further possible explanation for the observed difference in reported and corrected change: if they do not pay close attention to the experimental question wordings, interviewers may enter "yes" meaning that the situation is still the same, when in fact the question was worded as a CHANGED question. When the follow-up question is triggered, the interviewer may either backtrack and correct the answer to the proactive question, or simply fill in the follow-up question with the same answer category (see, Jäckle, Al Baghal and Eckman 2015).

In the IP7 panel of Figure 3, looking first at reported change (dark bars), we see that the CHANGED format collects more reports of change than the other formats. Statistically, the three other formats are not different from each other, and CHANGED is different from each of them

(Appendix Table 4). When we look instead at corrected change, we see less change in the CHANGED format, and now the differences between formats are no longer significant.

In the MC study, the pattern is different: the CHANGED format collects fewer reports of change than the STILL format; neither STILL nor CHANGED is significantly different from STILL/CHANGED at the 5% level. The new REMIND ASK format, however, collects significantly more reports of change than STILL, CHANGED and STILL/CHANGED. The independent question, which was not used in the IP studies, collects more reports of change than each of the dependent interviewing formats. This result is as found in previous research: simply asking the same questions wave after wave leads to overreports of change (e.g. Lynn and Sala 2006; Perales 2014). In the MC study, the difference between reported change and corrected change is not as pronounced as in the IP studies, which could be due to stricter supervision of telephone interviewers than face-to-face interviewers, leading to fewer interviewer errors.

What are the mechanisms through which the question wording affects reporting of change?

Tables 4 and 5 present results from the logistic models that test the satisficing, acquiescence and motivated misreporting mechanisms. For the moderators expected to increase differences in *satisficing* between formats (moderators 1, 2, 3, 4, 5, 6, 7, 8, 9, 13) there is only one significant interaction, in the MC data. The satisficing mechanisms predicts larger differences in the reporting of change between question formats for respondents with low cognitive ability compared to high ability. The direction of the observed effect is however the opposite: the difference in reporting change between formats increases with better working memory. For the moderators expected to increase differences in *acquiescence* between formats (10, 11, 12) none of the interactions are significant. For the variables used to test the *motivated*

misreporting mechanism (1, 2, 3, 4, 5, 13, 14) there are some significant main effects. Motivated misreporting predicts that respondents with higher cognitive ability and those more busy are less likely to report a change in all question formats. The main effects of working memory and busyness in the MC data are however in the opposite direction. The mechanisms also predicts that respondents are more likely to report change in all question formats, if others are present during the interview. In the IP3 data the effect of others present is in the expected direction, however the interaction with question format is significant, suggesting that the effect is not the same across question formats. We interpret these results as lack of support for any of the three mechanisms.

The hypothesis that biases in decision making heuristics may affect the reporting of change is not supported either. In separate logistic models the probability that a respondent mentioned something that had stayed the same (based on the coded open-ended responses) was no different between dependent interviewing formats in the MC data, and neither was the probability that they mentioned something that had changed or would change in future (results not shown). Thus there is no evidence that asking about change triggers thoughts of change, and vice versa.

Which question wording produces the most accurate measures of change?

The MC survey offers a link to administrative records, with which we can check the accuracy of reports of change to address Research Question 3. For each respondent, we derived indicators of change in unemployment benefits receipt, income support receipt, and employment status, from both the survey data and the administrative records. (Because there is almost no difference

between the reported and corrected change in the MC study (see Figure 3), we focus on reported change in the survey data.)

The survey and administrative indicators of change agree in 85.4% of the items (standard error: 0.66% points), and there are no significant differences between the question formats (tested with a Wald test accounting for clustering of items within respondents). Surprisingly, the independent question works just as well as all dependent interviewing formats. However, this aggregate comparison of the change indicator masks some differences in over- and underreporting of change. We therefore examine different types of transitions in more detail, by developing transition variables for each of the five possible states (receipt of unemployment benefits, receipt of income support, full-time employment, part-time employment and mini job) from the two sources (survey responses and administrative data).

Figure 4 shows the relationship between the transition indicators from the survey and the administrative records for all five variables together, by question format. The rows show the transition status according to the administrative records, the columns the transition according to the survey reports. The five different bars in each cell correspond to the five question formats and the height is the row percent.

The cells on the main diagonal ("No-No"…"Yes-Yes") show items where the survey transition status matches the records. The cells on the counter diagonal show items where the status in the survey was misreported in both waves: a rare event. The first row of the figure corresponds to "No-No" in the administrative data: according to the records, these states did not apply in either wave 1 or 2. Most of the survey reports in this row were also "No-No": 94-95% of the survey reports agreed with the administrative records in both waves: respondents are quite accurate when reporting that a state does not apply in either wave, regardless of question format

(Pearson chi-square test, corrected for the clustering of items in respondents and converted into an F statistic (Rao and Scott 1984), to test the correlation between survey reports and format: F(11.9, 13406.2)=0.85; p= 0.59). There are only short bars in the rest of the top row, indicating rare misreports in wave 2 ("No-Yes"), in wave 1 ("Yes-No") or in both waves ("Yes-Yes").

In the second row, there is more error. The true status for all items in this row, according to the administrative data, is "No-Yes" – transition into a state such as income support receipt or part-time employment. In this row, the most frequent survey report in all formats was "Yes-Yes," which is an underreporting of true change. Although there are larger differences among the formats in this row, they are not significant (F(11.9, 2889.5)=1.28; p=0.22).

The third row corresponds to "Yes-No" in the administrative data – transition out of a state. Here the majority of survey reports were correct in all formats, but "No-No" was also reported by more than 30% of respondents, which again represents an underreporting of change. There are no significant differences by format (F(11.9, 2957.2)= 0.74; p= 0.71).

The bottom row of the figure contains items where the state applied in both waves ("Yes-Yes"). More than 70% of all respondents in every format gave the correct response in each wave. Here we do see significant differences between the formats (F(11.9, 8187.1)=1.88; p = 0.033), with STILL and CHANGED collecting more correct "Yes-Yes" reports than the independent format.

In sum, the formats tend to work similarly, except when the state applies in each wave, then the STILL and CHANGED formats work best. The independent format is susceptible to overreporting of change when the true state is "Yes-Yes," but not when it is "No-No". With independent interviewing, respondents are less likely to report continued receipt of an income source (for example) correctly, and instead appear to have transitioned off. Similar results have

been found in other studies showing that dependent interviewing reduces downward biases in spell durations (Jäckle 2008).

The errors in transition status do not appear to be driven by errors in the wave 1 response. The four cells in the northeast quadrant of the figure are those where respondents overreported in Wave 1 (reported "yes" when the true answer was "no") and those in the southwest quadrant are those were respondents underreported in wave 1. We note that errors in the reporting of transitions into or out of a state (the "No-Yes" and "Yes-No" rows) occur most often when the wave 1 status is incorrect; however, there are no differences between the dependent formats (which mention the wave 1 report) and independent interviewing (which does not). In addition, the error rates in the reporting of a continuing true state ("Yes-Yes") are highest with independent interviewing, which cannot be due to errors in the preloaded wave 1 responses, since respondents in the independent format are not reminded of their Wave 1 response.

6 Discussion

This study provides new evidence on how best to word proactive dependent interviewing questions to measure change and stability: using data from three experimental studies and linked administrative records, we show which question wordings perform best and that respondents do not exploit opportunities for low-effort response strategies given by reminding them of previous answers.

As in previous studies, our results show that proactive dependent interviewing reduces reporting of change (e.g. Lynn et al. 2006; Lynn and Sala 2006; Perales 2014) and produces more accurate measures of change than independent interviewing (Lynn et al. 2012). However the way in which proactive dependent interviewing questions are worded clearly matters. Our main findings are first, that the STILL version, the STILL/CHANGED and the

CHANGED/STILL versions work similarly, and second, that the REMIND, ASK version produces higher rates of change, though still not as high as independent interviewing. Interestingly, the CHANGED version produces higher rates of change than the STILL version in the IP studies, but lower rates in the MC study, a point we return to below. The comparison with administrative records shows that the STILL and CHANGED formats produce the most accurate measures of change.

We find no evidence suggesting that respondents exploit the opportunities for satisficing, acquiescence or motivated misreporting offered by reminding respondents of their answers from previous interviews. We also find no evidence that the wording influences decision making heuristics.

Taking together all the evidence from this article, as well as findings from previous research, we believe that the STILL format provides the best data quality and recommend that panel surveys use this wording. It reduces the overreporting of change seen with the independent question while not encouraging satisfying, acquiescence or motivated misreporting. Although the CHANGED format was just as accurate in our analysis of Research Question 3, we have some reservations about recommending this format: in the IP studies, a number of respondents said "yes" their status had changed, but subsequently gave the same answer in the follow up question as they had in the previous year. This effects is seen in Figure 3 as the difference between the dark and light bars. Although this effect occurred in the IP studies in all formats, it happened most often in the CHANGED format. Additional research is needed to understand what is behind that result, but it does suggest that the STILL format is a better choice.

One remaining puzzle in our results is why the CHANGED format collects more reports of change in the IP studies but fewer in the MC study. There are several differences between

these surveys that might explain the inconsistent results. As mentioned above, the IP studies are face-to-face, which allows for less supervision of interviewers than the telephone MC study. The topics also differed between the surveys, along with the base rates of change. The IP studies both included a small number of other non-experimental DI questions, most of which used variants of the STILL format: it is possible that interviewers may not have paid sufficient attention to the different question wordings and incorrectly used the format they were most familiar with (Jäckle, Al Baghal and Eckman 2015). It is also possible that cultural differences between the German and English respondents affect our findings. Any one of these factors could drive the differing outcomes we see in Figure 3, and more research is needed to understand these results thoroughly.

References

Al Baghal, Tarek. 2016. "Last Year Your Answer Was… The Impact of Dependent Interviewing Wording and Survey Factors on Reporting of Change." *Field Methods*:preprint online publication 1525822X16645073.

Aquilino, William S. 1993. "Effects of Spouse Presence During the Interview on Survey Responses Concerning Marriage." *Public Opinion Quarterly* 57(3):358-76.

Aughinbaugh, Alison, and Rosella M. Gardecki. 2008. "Assessing Acquiescence among Participants in the National Longitudinal Survey of Youth 1997." in *Unpublished manuscript*. Washington DC: Bureau of Labour Statistics.

Baron, Reuben M, and David A Kenny. 1986. "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):1173.

Barry, Danielle, Marsha E. Bates, and Erich Labouvie. 2008. "FAS and CFL Forms of Verbal Fluency Differ in Difficulty: A Meta-analytic Study." *Applied Neuropsychology* 15(2):97-106.

Conrad, Frederick G. , Lance J. Rips, and Scott S. Fricker. 2009. "Seam Effects in Quantitative Responses." *Journal of Official Statistics* 25(3):339-61.

Corti, Louise, and Pamela Campanelli. 1992. "The Utility of Feeding Forward Earlier Wave Data for Panel Studies." Pp. 109-18 in *Survey and Statistical Computing*, edited by Andrew Westlake, Randy Banks, Clive Payne, and Terry Orchard. London: Elsevier Science Publishers.

Couch, Arthur, and Kenneth Keniston. 1960. "Yeasayers and naysayers: Agreeing response set as a personality variable." *The Journal of Abnormal and Social Psychology* 60(2):151-74.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein, and Kaiya Liu. 2011. "Partitioning Default Effects: Why People Choose Not to Choose, ." *Journal of Experimental Psychology: Applied* 17(4):332-41.

Eckman, Stephanie, and Frauke Kreuter. 2015. "Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden." in *IAB-Discussion Paper 29/2015*. Nürnberg IAB.

Eckman, Stephanie, Frauke Kreuter, Antje Kirchner, Annette Jäckle, Roger Tourangeau, and Stanley Presser. 2014. "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys." *Public Opinion Quarterly* 78(3):721-33.

Eggs, Johannes, and Annette Jäckle. 2015. "Dependent Interviewing and Sub-Optimal Responding." *Survey Research Methods* 9(1):15-29.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19(4):25-42.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. "A Very Brief Measure of the Big-Five Personality Domains." *Journal of Research in Personality* 37:504-28.

Hayman, Max. 1942. "Two Minute Clinical Test for Measurement of Intelectual Impairment in Psychiatric Disorders." *Archives of Neurology & Psychiatry* 47(3):454-64.

Hill, Daniel H. 1994. "The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey." *Journal of Official Statistics* 10(4):359-80.

Holbrook, Allyson L, Jon A Krosnick, David Moore, and Roger Tourangeau. 2007. "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly* 71(3):325-48.

Holmberg, Anders. 2004. "Pre-printing Effects in Official Statistics: An Experimental Study." *Journal of Official Statistics* 20(2):341-55.
Hoogendoorn, Adriaan W. 2004. "A Questionnaire Design for Dependent Interviewing that Addresses the Problem of Cognitive Satisficing." *Journal of Official Statistics* 20(2):219-32.
Hope, Steven, Pamela Campanelli, Gerry Nicolaas, Peter Lynn, and Annette Jäckle. 2014. "The Role of the Interviewer in Producing Mode Effects: Results from a Mixed Modes Experiment Comparing Face-to-Face, Telephone and Web Administration." in *ISER Working Paper 2014-20*. Colchester: University of Essex.
IAB. 2011. "Integrierte Erwerbsbiographien (IEB) V09.00." Nuremberg: Institute for Employment Research.
—. 2013. "Beschäftigtenhistorik (BeH), Version 09.03.00." Nuremberg: Institute for Employment Research.
Jäckle, Annette. 2008. "Measurement Error and Data Collection Methods: Effects on Estimates from Event History Data." in *ISER Working Paper*. Colchester: University of Essex.
—. 2009. "Dependent Interviewing: A Framework and Application to Current Research." Pp. 93-111 in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. Chichester: Wiley.
Jäckle, Annette , and Peter Lynn. 2007. "Dependent Interviewing and Seam Effects in Work History Data." *Journal of Official Statistics* 23(4):529-51.
Jäckle, Annette, Tarek Al Baghal, and Stephanie Eckman. 2015. "Interviewer and Respondent Behaviours when Measuring Change with Dependent Interviewing." in *Presented at the 6th conference of the European Survey Research Association*. Reykjavik.
Jäckle, Annette, Stephanie Eckman, and Austin Nichols. 2014. "How do Web Respondents React to Dependent Interviewing? Confidentiality Concerns and Reported Change Compared to Face-to-Face Administration." in *Presented at the 4th international Panel Survey Methods Workshop*. Ann Arbor, MI.
Jacobebbinghaus, Peter, and Stefan Seth. 2007. "The German Integrated Employment Biographies Sample IEBS." *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 127(2):335-42.
John, Oliver P., and Sanjay Srivastava. 1999. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." in *Handbook of Personality: Theory and Research*, edited by O. P. John and L. A. Pervin. New York: Guilford Press.
Knowles, Eric S., and Christopher A. Condon. 1999. "Why People Say "Yes": A Dual-Process Theory of Acquiescence." *Journal of Personality and Social Psychology* 77(2):379-86.
Knowles, Eric S., and K. T. Nathan. 1997. "Acquiescent Responding in Self-Reports: Cognitive Style or Social Concern?" *Journal of Research in Personality* 31(2):293-301.
Köhler, Markus, and Ulrich Thomsen. 2009. "Data Integration and Consilidation of Administrative Data from Various Sources. The Case of Germans Employment Histories." *Historical Social Research* 34(3):215-29.
Kreuter, Frauke, Susan McCulloch, Stanley Presser, and Roger Tourangeau. 2011. "The Effects of Asking Filter Questions in Interleafed Versus Grouped Format." *Sociological Methods & Research* 40(1):88-104.
Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-36.

Krosnick, Jon A., and Stanley Presser. 2010. "Questionnaire Design." Pp. 263-313 in *Handbook of Survey Research (Second Edition)*, edited by Peter V. Marsden and James D. Wright. West Yorkshire, England: Emerald Group.

Lavrakas, Paul J., Trevor N. Tompson, and Robert Benford. 2010. "Investigating Data Quality in Cell Phone Surveying." in *Presented at the 65th annual conference of the American Association for Public Opinion Research*. Chicago IL.

Lemaitre, Georges. 1992. "Dealing with the Seam Problem for the Survey of Labour and Income Dynamics." in *SLID Research Paper Series*. Ottawa: Statistics Canada.

Lugtig, Peter, and Annette Jäckle. 2011. "Can I Just Check…? Effects of Edit Check Questions on Measurement Error and Survey Estimates." in *ISER Working Paper*. Colchester: University of Essex.

Lynn, Peter. 2009. "Sample Design for Understanding Society." in *Understanding Society Working Paper 2009-01*. Colchester: University of Essex.

Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2006. "The Effects of Dependent Interviewing on Responses to Questions on Income Sources." *Journal of Official Statistics* 22(3):357-84.

Lynn, Peter, Annette Jäckle, Stephen P**.** Jenkins, and Emanuela Sala. 2012. "The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study." *Journal of the Royal Statistical Society, Series A* 175(1):289-308.

Lynn, Peter, and Emanuela Sala. 2006. "Measuring Change in Employment Characteristics: the Effects of Dependent Interviewing." *International Journal of Public Opinion Research* 18(4):500-09.

Mathiowetz, Nancy A., and Katherine A. McGonagle. 2000. "An Assessment of the Current State of Dependent Interviewing in Household Surveys." *Journal of Official Statistics* 16(4):401-18.

Moore, Jeffrey, Nancy Bates, Joanne Pascale, and Aniekan Okon. 2009. "Tackling Seam Bias Through Questionnaire Design." Pp. 73-92 in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. Chichester: Wiley.

Murray, T. Scott, Sylvie Michaud, Mike Egan, and Georges Lemaitre. 1991. "Invisible Seams? The Experience with the Canadian Labour Market Activity Survey." Pp. 715-30 in *Proceedings of the US Bureau of the Census Annual Research Conference*. Washington DC: US Census Bureau.

Pascale, Joanne, and Alice McGee. 2008. "Using Behavior Coding to Evaluate the Effectiveness of Dependent Interviewing." *Survey Methodology* 34(2):143-51.

Perales, Francisco. 2014. "How Wrong Were We? Dependent Interviewing, Self-Reports and Measurement Error in Occupational Mobility in Panel Surveys." *Longitudinal and Life Course Studies* 4(3):299 – 316.

Rammstedt, Beatrice, Lewis R. Goldberg, and Ingwer Borg. 2010. "The Measurement Equivalence of Big-Five Factor Markers for Persons with Different Levels of Education." *Journal of Research in Personality* 44:53-61.

Rao, Jon NK, and Alastair J Scott. 1984. "On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data." *The Annals of Statistics*:46-60.

Rips, Lance J., Frederick G. Conrad, and Scott S. Fricker. 2003. "Straightening the Seam Effect in Panel Surveys." *Public Opinion Quarterly* 67(4):522-54.

Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers." *American psychologist* 54(2):93.
Silver, Brian D, Barbara A Anderson, and Paul R Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(02):613-24.
The American Association for Public Opinion Research. 2016. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition." AAPOR.
Toplak, Maggie E., Richard F. West, and Keith E. Stanovich. 2011. "The Cognitive Reflection Test as a Predictor of Performance on Heuristics-and-Biases Tasks." *Memory & Cognition* 39:1275-89.
Tourangeau, Roger, Frauke Kreuter, and Stephanie Eckman. 2012. "Motivated Underreporting in Screening Interviews." *Public Opinion Quarterly* 76(3):453-69.
—. 2015. "Motivated Misreporting: Shaping Answers to Reduce Survey Burden." Pp. 24-41 in *Survey Measurement: Techniques and Findings from Recent Research.*, edited by Uwe Engel: Campus.

Figure 1: Distribution of Moderator Variables in IP3 Study

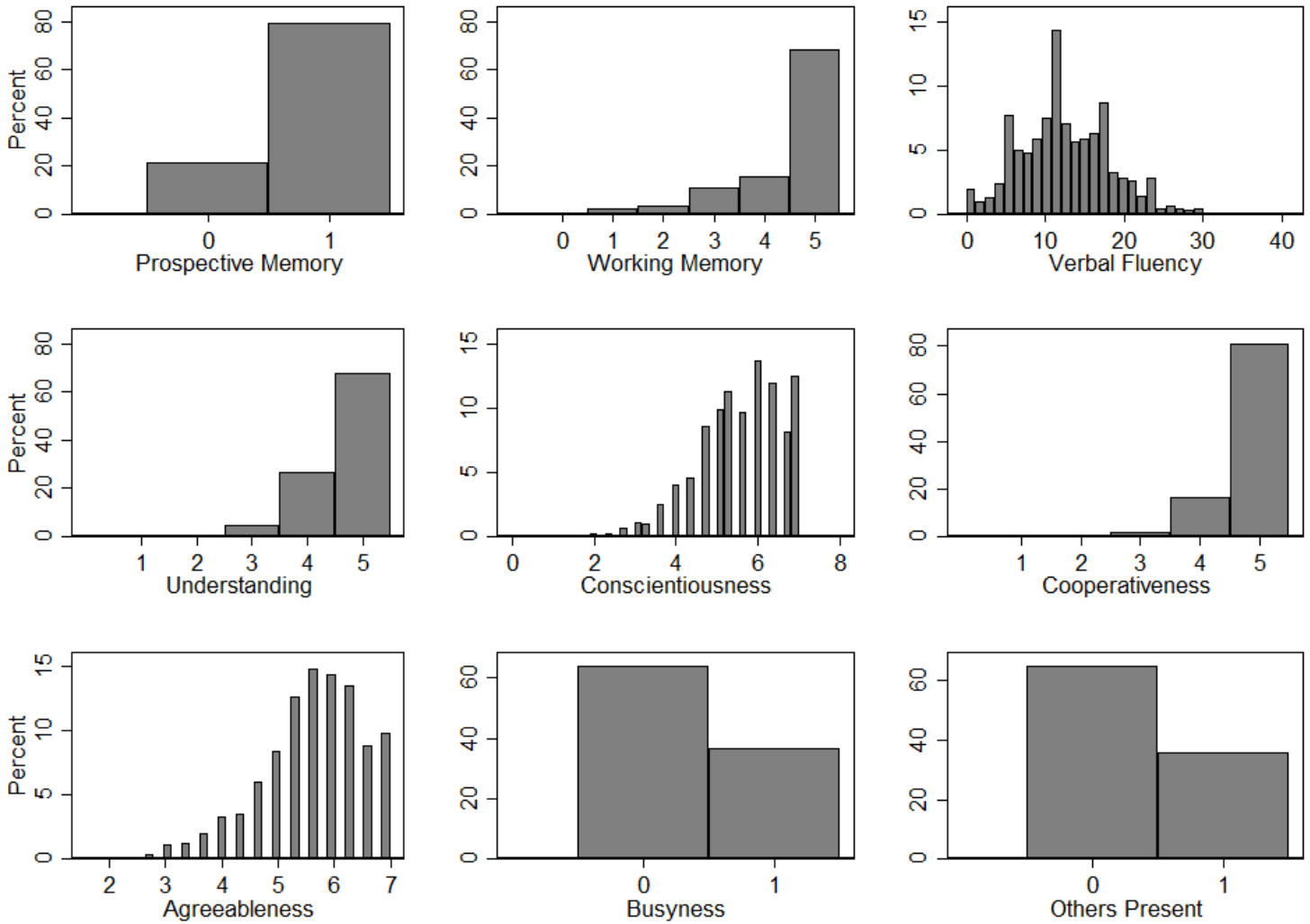

Figure 2: Distribution of Moderator Variables in MC Study

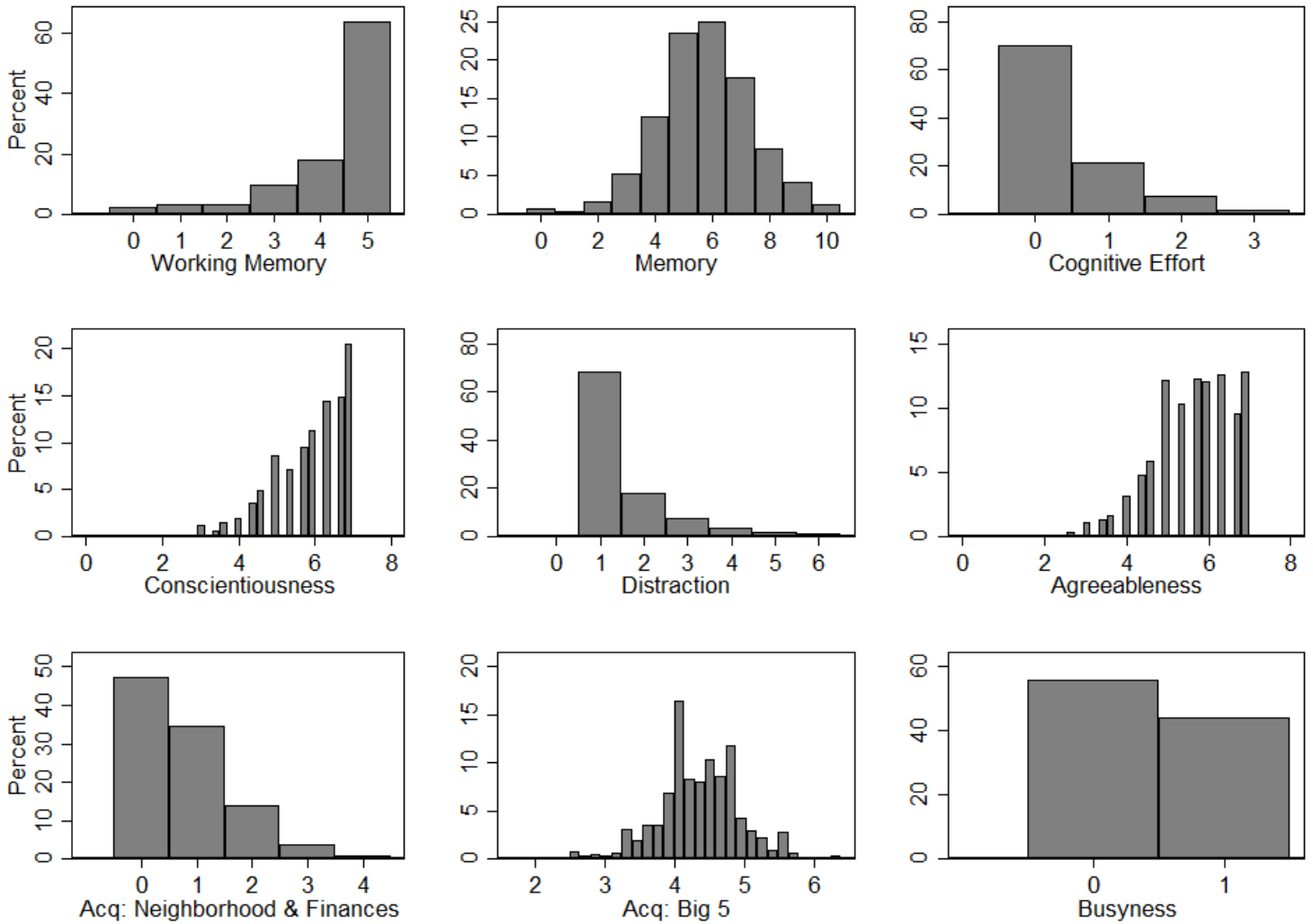

Figure 3: Change Reporting by Format

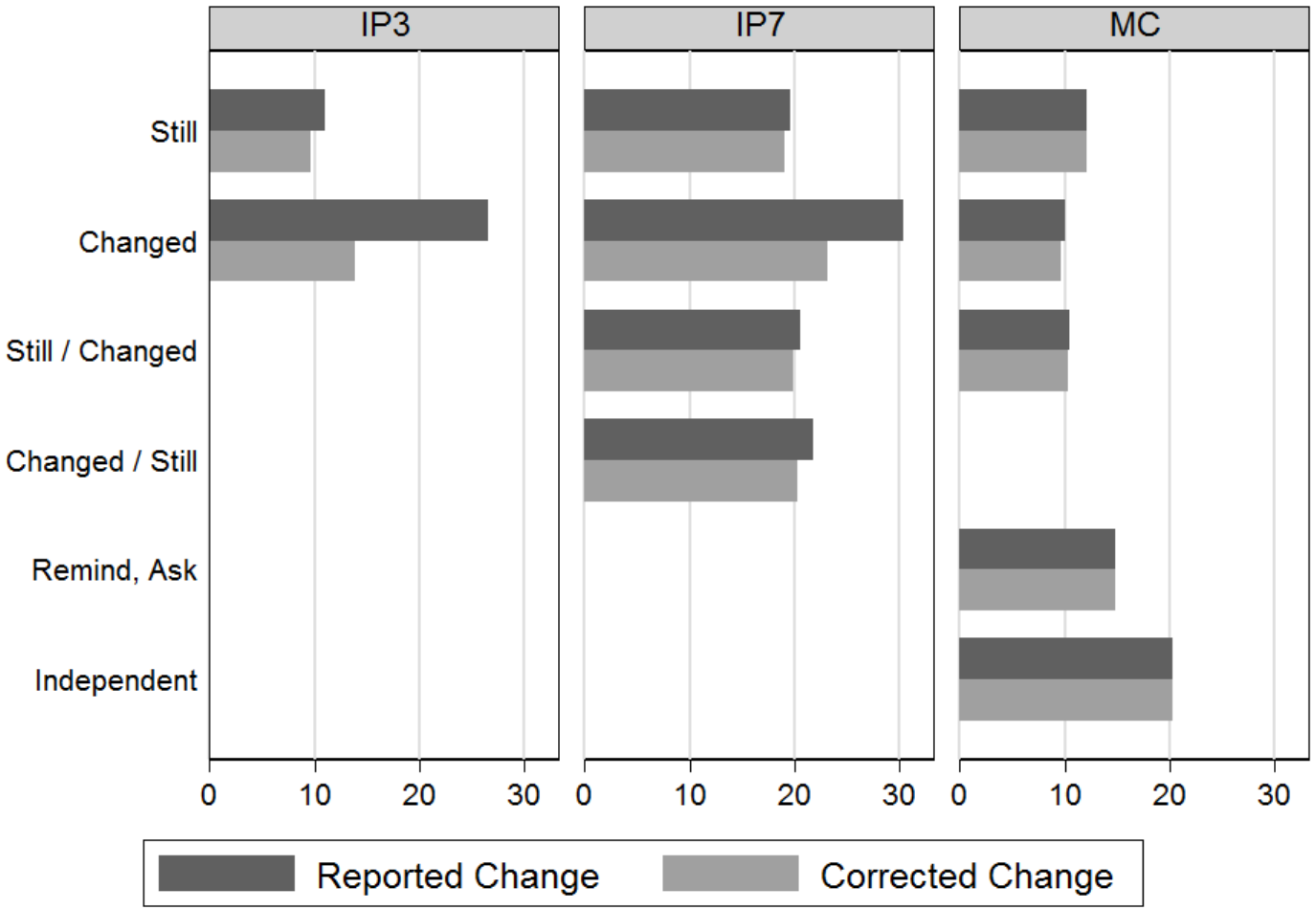

Figure 4: Wave 1 and 2 Reports in MC Survey, by Format and Administrative Data in Wave 1 and 2

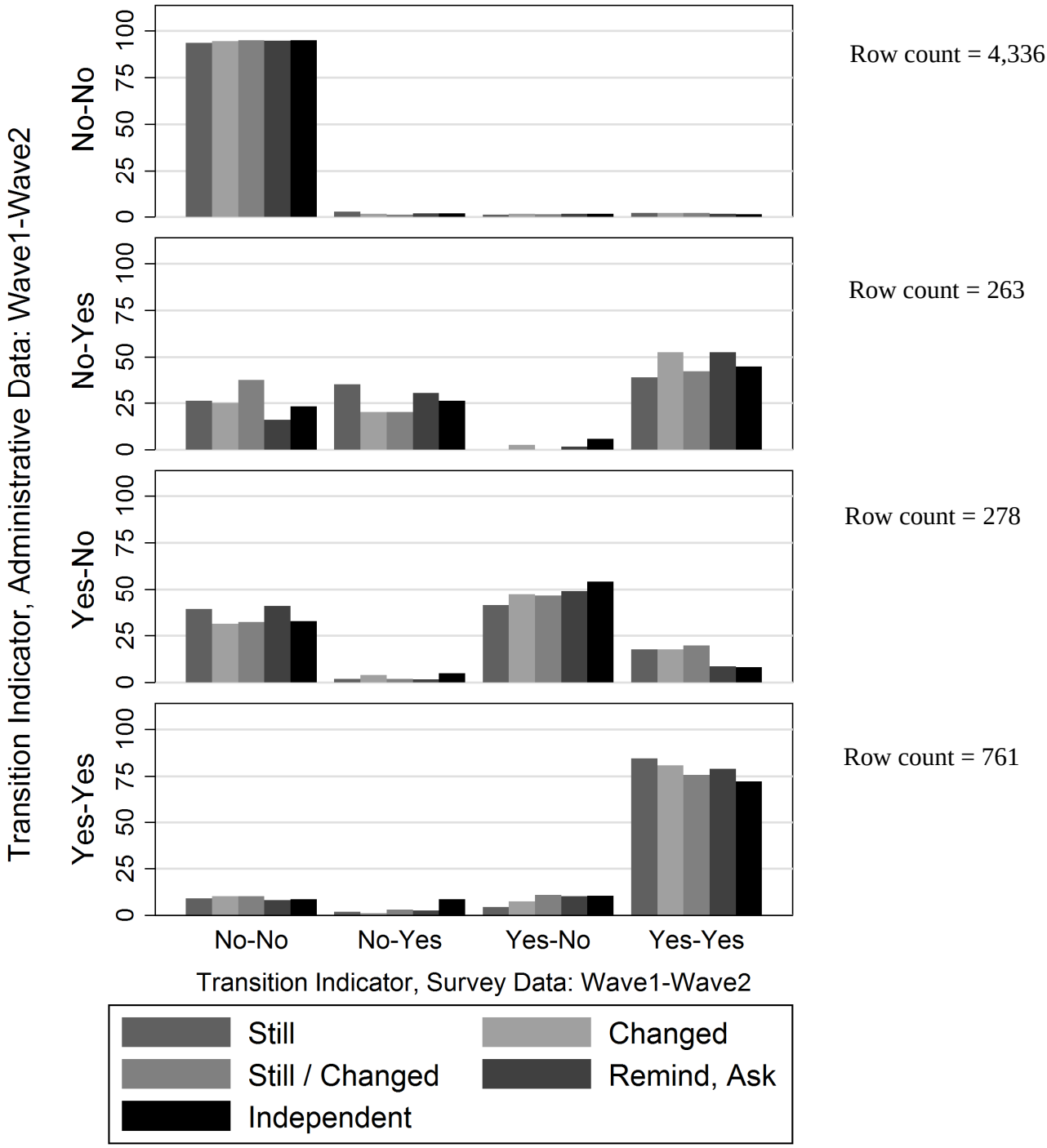

Table 1: Research Questions and Studies

Study	Research Question 1: Does the wording of proactive dependent interviewing questions affect the likelihood of reporting a change?	Research Question 2: What are the mechanisms through which the question wording affects reporting of change?	Research Question 3: Which question wording produces the most accurate measures of change?
IP3	X	X	
IP7	X		
MC	X	X	X

Table 2: Experimental Variation in Dependent Question Wording, by Survey

Format	Wording	Response Options	IP3	IP7	MC
STILL	"Is this still the case?"	YES NO	X	X	X
CHANGED	"Has this changed?"	YES NO	X	X	X
STILL/CHANGED	"Is this still the case or has it changed?"	STILL CHANGED		X	X
CHANGED/STILL	"Has that changed or is it still the case?"	CHANGED STILL		X	
REMIND/ASK	Reminder followed by independent question as asked in wave 1				X
INDEPENDENT	Independent question as asked in wave 1				X

Table 3: Moderator Variables in IP3 and MC Studies

Cognitive ability: We focused on aspects of memory which we expect to correlate with more accurate reporting of change: using the reminder as a prompt, the respondent has to recall what their situation at the previous interview was, and compare that with their current situation to decide whether or not there has been a change. We used several tests as well as interviewer assessments of how well respondents understood questions as proxies for cognitive ability:					
	Items	Description	Coding	IP3[a]	MC[b]
1	*Prospective Memory*: the ability to remember to do something	Respondent told s/he would receive a piece of paper and a pencil at some point and that s/he should then write her/his date of birth in the top left hand corner. If the respondent did not write the date of birth within five seconds of being handed the paper, the interviewer prompted: *"You were going to do something when I gave you the paper and pencil. Can you remember what it was?"*	1 if the respondent wrote the date in the top left-hand corner without being prompted, 0 otherwise	X	
2	*Working Memory*: the short-term processes used to store and make use of information for complex tasks	Respondent asked to subtract 7 from 100, to subtract 7 from the resulting answer, etc, for a total of five times. The interviewer recorded each of the five answers. (Hayman 1942)	Number of correct subtractions	X	X
3	Phonemic *Verbal Fluency*: the ability to generate words starting with the same letter (related to executive function, which is important in overriding more automatic responses)	Respondent randomly assigned to the letters F, A, or S and asked to say as many words as s/he could think of in one minute that started with that letter. (Barry, Bates and Labouvie 2008)	Number of correct words	X	
4	*Memory*: the ability to remember a list of words	Immediate word recall test: interviewer read out a list of 10 words that respondent was asked to recall in any order.	Number of words recalled accurately		X
5	Respondent's	Interviewer observation at the end of the interview: *"In*	1 if the rating was	X	

	Understanding of Questions	*general, how would you describe the respondent's understanding of the questions?"* (excellent, good, fair, poor, very poor)	excellent, 0 otherwise		

Cognitive effort: The willingness to expend cognitive effort was assessed using three quiz questions that measured whether respondents gave a (false) intuitive top-of-the head response or made the effort to query the intuitive answer and figure out the correct answer. This test provides a direct observation of satisficing that respondents may use with proactive dependent interviewing questions. In addition, we used a 15-item version of the Big 5 inventory of personality traits to measure conscientiousness, which refers to self-control, task-orientation, and rule-abiding, and could be predictive of cognitive effort, and interviewer observations about how cooperative the respondent was during the interview.

	Items	Description	Coding	IP3[a]	MC[b]
6	*Cognitive Effort*	Cognitive Reflection Test: Three quiz questions where the intuitive top-of-the-head response is false and figuring out the correct answer requires deliberate cognitive effort that over-rides intuitive low-effort decision making. The test provides a direct observation of "miserly cognitive processing," as opposed to self-reports on Need for Cognition, and is to some extent independent of cognitive ability (Frederick 2005; Toplak, West and Stanovich 2011).	1 if the respondent answered at least one of the three questions correctly, 0 otherwise.		X
7	Big 5 *Conscientiousness*	15-item version of the Big 5 inventory of questions that are designed to measure the traits of Extroversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (Gosling, Rentfrow and Swann 2003; John and Srivastava 1999)	We reverse coded items that measured the opposite of the trait and then derived mean scores by averaging over the three items related to the trait.	X	X
8	Respondent's *Cooperativeness*	Interviewer observation at the end of the interview: *"In general, the respondent's cooperation during the interview was…"* (very good, good, fair, poor, very poor).	1 if the rating was very good, 0 otherwise	X	

Cognitive distraction was measured by asking respondents which other activities they were engaged in during the CATI interview. Multi-tasking would have meant that the respondent concentrated less on listening to the questions and thinking of answers, resulting in less cognitive effort.

	Items	Description	Coding	IP3[a]	MC[b]
9	*Distraction*	List of 16 activities (plus "other" option) that respondent might have been engaged in during the interview. Respondent is asked to indicate all that apply. (Lavrakas, Tompson and Benford 2010)	1 if the respondent reported two or more activities, 0 otherwise.		X
10	Big 5 *Agreeableness*	15-item version of the Big 5 inventory of questions that are designed to measure the traits of Extroversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (Gosling, Rentfrow and Swann 2003; John and Srivastava 1999)	We reverse coded items that measured the opposite of the trait and then derived mean scores by averaging over the three items related to the trait.	X	X
11	*Acquiescence 1*	Two pairs of opposing items using 5-point agree-disagree scales about the respondent's neighbourhood, and two opposing pairs and about how they make financial decisions. If respondent answered these questions consistently, answers should be symmetrical about the middle category (3). If respondent acquiesced, the expected mean for each item pair is larger than three. (Hope et al. 2014)	Score from 0 to 4 indicating the number of pairs where respondents gave contradictory answers.		X
12	*Acquiescence 2*	The Big 5 personality traits are measured with 15 items using 7-point agree-disagree response scales. For four of the items there is an opposing statement. If the respondent answered these questions consistently, answers should be symmetrical about the middle category (4). If the respondent acquiesced the expected mean for each item pair is larger than four. (Rammstedt, Goldberg and Borg 2010)	1 if mean score across the 8 items > 4, 0 otherwise		X
13	*Busyness*	Whether the respondent had young children aged five or younger in the household, whether s/he worked more than 37.5 hours a week, or commuted more than 60 minutes to work.	1 if the respondent met any of the three criteria, and 0 otherwise.	X	X
14	*Others Present*	Interviewer observation at the end of the interview	Yes/No	X	
15	Open questions	Follow-ups to four dependent interviewing questions asking respondents what went through their mind when thinking	'No change' indicator coded as 1 if answer		X

		about how their current situation compared to the previous year. See Appendix Table 5 for the coding frame. (Dinner et al. 2011)	referred to something that had not changed, 0 otherwise. 'Change' indicator coded as 1 if answer referred to something that had changed/would change in future, 0 otherwise.		

[a] The full wording of questions can be found in the IP3 questionnaire at

https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires.

[b] Full wording of MC questions can be found in Appendix Table 6.

Table 4: Probability of Reporting Change, IP3 Study

Model	Main Effects				Interaction		N	Pseudo R^2	LL
	Variable		Changed (vs Still)						
1 prospective memory	-0.352		0.938	***	0.116		2,566	0.0425	-1177.1
	(0.216)		(0.231)		(0.265)				
2 working memory	0.066		1.038		0.005		2,520	0.0418	-1156.0
	(0.108)		(0.610)		(0.134)				
3 verbal fluency	0.022		1.460	***	-0.033		2,565	0.0411	-1178.6
	(0.015)		(0.278)		(0.019)				
5 understanding of questions	0.066		1.218	***	-0.249		2,568	0.0419	-1178.1
	(0.189)		(0.192)		(0.237)				
7 conscientiousness	-0.094		0.916		0.030		2,250	0.0439	-1011.1
	(0.091)		(0.647)		(0.115)				
8 cooperativeness	0.045		1.246	***	-0.243		2,568	0.0418	-1178.3
	(0.221)		(0.243)		(0.274)				
10 agreeableness	0.047		0.922		0.025		2,247	0.0433	-1011.2
	(0.099)		(0.717)		(0.1.42)				
13 busyness	-0.121		0.982	***	0.190		2,577	0.0412	-1185.0
	(0.191)		(0.137)		(0.238)				
14 others present	0.440	*	1.256	***	-0.526	*	2,568	0.0438	-1175.8
	(0.180)		(0.147)		(0.230)				

Coefficients and standard errors in brackets, from separate logistic models. Standard errors adjusted for clustering of items within respondents. *** P<0.001, ** 0.001≤P<0.01, * 0.01≤P<0.05

Table 5: Probability of Reporting Change, MC Study

	Main Effects								
Model	Variable		Changed (vs Still)		Interaction		N	Pseudo R^2	LL
2 working memory	0.152	**	-1.005	**	0.187	*	6,227	0.005	-2157.5
	(0.057)		(0.372)		(0.082)				
4 memory	0.022		-0.338		0.021		6,227	0.002	-2163.8
	(0.042)		(0.354)		(0.059)				
6 cognitive effort	-0.276		-0.303	*	0.288		6,224	0.003	-2162.1
	(0.144)		(0.119)		(0.204)				
7 conscientiousness	0.007		0.052		-0.447		6,227	0.002	-2164.2
	(0.070)		(0.604)		(0.102)				
9 distraction	0.378	*	-0.184		-0.088		6,227	0.005	-2156.7
	(0.150)		(0.118)		(0.207)				
10 agreeableness	-0.064		-0.784		0.100		6,214	0.002	-2160.2
	(0.082)		(0.649)		(0.113)				
11 acquiescence 1	-0.073		-0.313	*	0.128		6,227	0.002	-2163.6
	(0.074)		(0.128)		(0.104)				
12 acquiescence 2	0.197		-0.045		-0.236		6,227	0.002	-2163.4
	(0.161)		(0.185)		(0.217)				
13 busyness	0.385	**	-0.129		-0.160		6,227	0.005	-2156.8
	(0.138)		(0.132)		(0.193)				

Coefficients and standard errors in brackets, from separate logistic models. Standard errors adjusted for clustering of items within respondents. *** P<0.001, ** 0.001≤P<0.01, * 0.01≤P<0.05

Appendix

Appendix Table 1: Dependent Interviewing Questions in IP3 Survey

Variable	Version	Question text
General health	Independent	In general, would you say your health is... [Excellent, Very good, Good, Fair, Poor]
	DI reminder	The last time we interviewed you on *[ff_IntDate]* , you said that, in general, your health was *[ff_sf1]*.
Permanent job	Independent	Leaving aside your own personal intentions and circumstances, is your job... [A permanent job, Or is there some way that it is not permanent?]
	DI reminder	Last time we interviewed you on *[ff_IntDate]* you said that, leaving aside your own personal intentions and circumstances, your job was a permanent job *{if ff_jbterm1 = 1}* / was not a permanent job in some way *{if ff_jbterm1 = 2}*.
Working hours (employees)	Independent	Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week? [Number]
	DI reminder	Last time we interviewed you, you said that in your (main) job, you were expected to work *[ff_jbhrs]* hours in a normal week, excluding overtime and meal breaks.
Working hours (self-employed)	Independent	How many hours in total do you usually work in a week in your job? [Number]
	DI reminder	Last time we interviewed you, you said that you usually work *[ff_jshrs]* hours, in total each week, in your job.

Appendix Table 2: Dependent Interviewing Questions in IP7 Survey

Variable	Version	Question text
Number of bedrooms	Independent	How many bedrooms are there here excluding any bedrooms you may let or sublet? [Number]
	Reminder	When we interviewed you on *[ff_Idate]*, you said you had *[ff_HsBeds]* bedroom(s), excluding any you may sublet and *[ff_HsRooms]* other rooms, excluding kitchens and bathrooms.
Housing tenure	Independent	Does your household own this accommodation outright, is it being bought with a mortgage, is it rented or does it come rent-free? [Owned outright, Owned/being bought on mortgage, Shared ownership (part-owned part-rented), Rented, Rent free, Other]
	Reminder	Last time you said that this accommodation was *[ff_HsOwnd]*.
Mortgage payment	Independent	How much was your last total monthly instalment on all mortgages or loans for this property? [Number]
	Reminder	Last time your total monthly instalment on all mortgages or loans for this property was *[ff_xpmg]*.
Rent payment	Independent	How much was the last rent payment, including any services or water charges but after any rebates? [Number]
	Reminder	Last time you paid *[ff_rent] [ff_rentwc]*.
Stay or move	Independent	If you could choose, would you stay here in your present home or would you prefer to move somewhere else? [Stay here, Prefer to move]
	Reminder	Last time we interviewed you on *[ff_IntDate]* you said that if you could choose, you would stay here in your present home *{if ff_lkmove = 1}*/ prefer to move somewhere else *{if ff_lkmove = 2}*.
Education	Independent	Are you... [At School, At Sixth Form College, At Further Education (FE) College, At Higher Education (HE) College, or at University?]
	Reminder	The last time we interviewed you on *[ff_IntDate]* you said that you were *[ff_edtype]*.
Permanent job	Independent	Leaving aside your own personal intentions and circumstances, is your job... [A permanent job, Or is there some way that it is not permanent?]
	Reminder	Last time we interviewed you on *[ff_IntDate]* you said that, leaving aside your own personal intentions and circumstances, your job was a permanent job *{if ff_jbterm1 = 1}* / was not a permanent job in some way *{if ff_jbterm1 = 2}*.
Industry	Independent	What does the firm/organisation you work for mainly make or do at the place where you work? [Text]
	Reminder	Last time you said that the firm or organisation where you work, makes or does *[ff_jbsic07]*.

Occupation	Independent	What was your main job last week? Please tell me the exact job title and describe fully the sort of work you do. [Text]
	Reminder	Last time you described your occupation in your **main** job as *[ff_jbsoc00]*.
Employment status	Independent	Are you an employee or self-employed? [Employee, Self-employed]
	Reminder	Last time you said that you were an employee *{if ff_jbsemp = 1}* / self-employed *{if ff_jbsemp = 2}*.
Firm size	Independent	How many people are employed at the place where you work? [1 – 2, 3 – 9, 10 – 24, 25 – 49, 50 – 99, 100 – 199, 200 – 499, 500 – 999, 1000 or more, Don't know but fewer than 25, Don't know but 25 or more]
	Reminder	Last time, you said that there were *[ff_jbsize]* people employed at the place you work.
Working hours (employees)	Independent	Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week? [Number]
	Reminder	Last time we interviewed you, you said that in your **main** job, you were expected to work *[ff_jbhrs]* hours in a normal week, excluding overtime and meal breaks.
Gross pay	Independent	What was your most recent gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance or pension contributions, union dues and so on? [Number]
	Reminder	When we interviewed you on *[ff_IntDate]* , you said that last time you were paid, your gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance or pension contributions, union dues and so on - was £*[ff_paygl] [ff_paygwc]*.
Net pay	Independent	What was your most recent take home pay, that is after any deductions were made for tax, National Insurance, pensions, union dues and so on? [Number]
	Reminder	And when we interviewed you on *[ff_IntDate]*, you said that last time you were paid, your net pay - that is after any deductions were made for tax, National Insurance, pensions, union dues and so on - was £ *[ff_paynl] [ff_paynwc]*.
Pay type	Independent	How is your pay calculated, in particular are you salaried or paid by the hour? [Salaried, Basic salary plus commission, Paid by the hour, Other]
	Reminder	Last time you said that you were salaried *{if ff_paytyp = 1}* / you received a basic salary plus commission *{if ff_paytyp = 2}* / you were paid by the hour *{if ff_paytyp = 3}*.
Travel to work	Independent	And how do you usually get to your place of work? [Drive myself by car or van, Get a lift with

(employees)		someone from household, Get a lift with someone outside the household, Motorcycle/moped/scooter, Taxi/minicab, Bus/coach, Train, Underground/Metro/Tram/Light railway, Cycle, Walk, Other]
	Reminder	Last time you said that you usually travel to work by [ff_worktrav] *{if ff_worktrav = 4|5|6|7|8}* / [ff_worktrav] to work *{if ff_worktrav =1|2|3|9|10}*.
Working hours (self-employed)	Independent	How many hours in total do you usually work in a week in your job? [Number]
	Reminder	Last time we interviewed you, you said that you usually work *[ff_jshrs]* hours, in total each week, in your job.
Self-employment partnership	Independent	Are you working on your own account or are you in partnership with someone else? [Own account (sole owner), In partnership]
	Reminder	Last time we interviewed you, you said that you were working on your own account (sole owner) *{if ff_jspart = 1}* / in partnership with someone else *{if ff_jspart = 2}*.
Travel to work (self-employed)	Independent	And how do you usually get to your place of work? [Drive myself by car or van, Get a lift with someone from household, Get a lift with someone outside the household, Motorcycle/moped/scooter, Taxi/minicab, Bus/coach, Train, Underground/Metro/Tram/Light railway, Cycle, Walk, Other]
	Reminder	Last time you said that you usually travel to work by ['ff_jsworktrav'] *{if ff_jsworktrav = 1|4|5|6|7|8}* / travel to work by getting a ['ff_jsworktrav'] *{if ff_jsworktrav = 2|3}* / ['ff_jsworktrav'] to work *{if ff_jsworktrav = 9|10}*.

Appendix Table 3: Dependent Interviewing Questions in MC Survey

Variable	Version	Question text
Household Size	Independent	How many people live in your household permanently, yourself included? Please include all children living in your household.
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said that [HHMEMB] person(s) lived in your household permanently, including yourself and children.
Marital Status	Independent	What is your marital status? Are you … [Married and living with your spouse, Cohabiting in a (same sex) Civil Partnership, Separated, Single, Divorced, Widowed, Separated from a Civil Partner, Annulated Civil Partnership, Widowed Civil Partnership]
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said you were [MASTAT].
Education	Independent	What is your highest school qualification? [9 categories, including "none"]
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said you had [QUALIFICATION].
Employment status	Independent	What is your current employment situation? Are you… [Self-employed, Full-time employee, Part-time employee, Irregular employment or "mini job", Apprentice, Parental leave, Registered unemployed, Military service or Voluntary Social Year, Not working (including pupils and students, long-term sick and pensioners without earnings]
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said you were [ACTIVITY].
Working hours	Independent	And how many hours do you work per week, including regular overtime?
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said you work [HOURS] hours per week, including regular overtime.
Earnings	Independent	How much was your last monthly pay? Please enter your gross pay, that is, before deduction of taxes and social security contributions. Please do not include irregular payments, such as vacation or back pay. Please do include pay for overtime. For self-employed activities please enter your monthly profit before deduction of taxes.
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said your monthly gross pay was [EARNINGS] Euros.
Investment Income	Independent	Did you or any other member of your household receive any income from interest or dividends during 2011, for example from savings, stocks, equity funds or bonds?

	Reminder	Last time we interviewed you, you said that {IF INTEREST==0: no-one in your household had received income from interest or dividends during 2010, for example from savings, stocks, equity funds or bonds} {IF INTEREST==1: you or another member of your household had received income from interest or dividends during 2010, for example from savings, stocks, equity funds or bonds}.
Rental Income	Independent	In 2011: Did you or any member of your household receive income from renting or leasing property?
	Reminder	Last time we interviewed you, you sad that {IF RENTAL==0: no-one in your household had received income from renting or leasing property during 2010} {IF RENTAL==1: You or another member of your household had received income from renting or leasing property during 2010}.
Child Benefit	Independent	In 2011: Did you or any other member of your household receive child benefit?
	Reminder	Last time we interviewed you, you said that {IF CHBEN==0: no-one in your household had received child benefit during 2010} {IF CHBEN==1: You or another member of your household had received child benefit during 2010}.
Maternity pay	Independent	In 2011: Did you or another member of your household receive a paternity or maternity pay?
	Reminder	Last time we interviewed you, you said that {IF MATPAY==0: no-one in your household had received paternity or maternity pay during 2010} {IF MATPAY==1: You or another member of your household had received paternity or maternity pay during 2010}.
Unemployment benefit 2	Independent	In 2011: Did you or another member of your household receive unemployment benefit 2, also known as Hartz IV?
	Reminder	Last time we interviewed you, you said that {IF UNEMP2==0: no-one in your household had received unemployment benefit 2, also known as Hartz IV, during 2010} {IF UNEMP2==1: You or another member of your household had received unemployment benefit 2, also known as Hartz IV, during 2010}.
Unemployment benefit 1	Independent	In 2011: Did you or another member of your household receive unemployment benefit 1?
	Reminder	Last time we interviewed you, you said that {IF UNEMP1==0: no-one in your household had received unemployment benefit 1 during 2010} {IF UNEMP1==1: You or another member of your household had received unemployment benefit 1 during 2010}.
Savings	Independent	Did you regularly save a certain amount of money during the last 12 months?
	Reminder	Last time we interviewed you in [INTMONTH] 2011, you said that you had {IF AMOUNT==0: not} regularly saved a certain amount of money during the previous 12 months.

Appendix Table 4: Coefficients and Standard Errors from Logistic Models of Change Reports (RQ1)

Format	**IP 3**		**IP 7**		**MC**	
Dependent Variable	Reported Change	Actual Change	Reported Change	Actual Change	Reported Change	Actual Change
Still	reference	reference	reference	reference	reference	reference
Changed	1.069***	0.404**	0.585***	0.241	-0.216*	-0.243*
	(0.112)	(0.127)	(0.175)	(0.176)	(0.0972)	(0.0980)
Still/Changed	–	–	0.0616	0.0508	-0.163	-0.178
			(0.185)	(0.187)	(0.0929)	(0.0931)
Changed/Still	–	–	0.133	0.0700	–	–
			(0.170)	(0.174)		
Remind, Ask	–	–	–	–	0.236**	0.245**
					(0.0873)	(0.0873)
Independent	–	–	–	–	0.612***	0.621***
					(0.0855)	(0.0855)
N	2,571	2,571	1,821	1,821	15,868	15,868
R^2	0.0423	0.0058	0.00953	0.00139	0.0148	0.0159
LL	-1178.3	-920.9	-977.6	-926.1	-6239.9	-6196.8

Standard errors in parentheses. Adjusted for clustering of items within respondents. *** $P<0.001$, ** $0.001 \leq P<0.01$, * $0.01 \leq P<0.05$.

Appendix Table 5: Coding Frame for Open Ended Questions (MC)

Notes	Code	Description	Examples
	No thoughts	Respondent didn't have any thoughts.	Includes "no thoughts, my situation hasn't changed"; "don't know"
Code all that apply	No change	The statement refers to something that has not changed	"That I have been married to my wife for over 50 years."
	Change	The statement refers to something that has changed	Including if it is unclear whether the change was in the last year or earlier
	Future change	The statement refers to something that the respondent expects or hopes will change in future	"When my daughter goes to university she will leave home."
	Unclear	Unclear whether statement is about something that has changed or not changed / change does not apply	Includes statements where "change" does not apply, e.g. "am I meant to answer the question now?"
Code all that apply	Description	Description of respondent's situation	"That I now have more money but also have to work more"
	Emotion	Explicit expression of emotion / feelings	"I am content/happy…", "I hope…"
	Evaluation	Evaluation / judgement of respondent's situation	Respondent makes a judgement about something, e.g. "good/shame that….is the way it is", "I'd prefer to be earning more"
If none of previous 3:	Other	General evaluation or comment that is not about the respondent himself	Includes general statements about the state of the economy
	Consideration - question	Respondent wondered about the definition of the question and what to include in the answer	"Whether animals count."
	Not understood	Respondent did not understand the question	"Strange question"

Appendix Table 6: Questions in MC Survey used as Moderator Variables

#[a]	**Question Text**
2	And now a few calculations. What is one hundred minus seven? [Number] And minus 7? [Number] And minus 7? [Number] And minus 7? [Number] And minus 7? [Number]
4	Now I am going to read out a list of words. We made a long list intentionally, so that it will be difficult for anyone to remember all the words. Most people only remember a few words. Please listen carefully, because I cannot repeat the list. When I finished reading, I will ask you to repeat as many of these words as you can out loud. The order is not important. Did you understand everything? Butter, Arm, Letter, Queen, Card, Grass, Corner, Stone, Book, Floor Now please tell me all the words that you can remember. [Check all that apply]
6	For the next set of questions, you can take notes if you want to. Have you got a pen and paper, or do you want to get this? • A bat and a ball cost EUR 1.10 in total. The bat costs Euro 1.00 more than the ball. How much does the ball cost? [Cents] • If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? [Minutes] • In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? [Days]
7 10 12	The following are different characteristics a person may have. Some of these characteristics will probably apply perfectly to you and others not at all. For some you may be undecided. Please answer using the following scale. The number 1 means: does not apply to me at all. The number 7 means: applies to me perfectly. You can use the numbers between 1 and 7 to nuance your opinion. I see myself as someone who… • does a thorough job. • is talkative. • is sometimes rude to others.

	• is original, comes up with new ideas. • worries a lot. • has a forgiving nature. • tends to be lazy. • is outgoing, sociable. • values artistic, aesthetic experiences. • gets nervous easily. • does things efficiently. • is reserved. • is considerate and kind to almost everyone. • has an active imagination. • is relaxed, handles stress well.
9	Were you engaged in any other activities during this interview? Did you … [Check all that apply: Watch TV, Drive a car, Sit at the computer (including checking emails), Do housework, Walk around, Eat or drink, Cook, Look after a child, Play, Help someone (e.g. a child), Speak with someone else, Send text messages, Change clothes, Practice personal care or grooming, Read something, Repair something, Or do something else?]
11	The next questions are about your neighborhood. To what extent do you agree with the following statements? [Strongly agree, Agree, Neither nor, Disagree, Strongly disagree] • This neighborhood is a good place to live. • Compared to other neighborhoods, this neighbourhood has more properties that are in a poor state of repair. • Compared to other neighborhoods, this neighbourhood has more problems such as litter, dog mess and graffiti. Compared to other neighborhoods, this neighborhood has more properties that are well kept.

11	The next questions are about how you make financial decisions. To what extent do you agree with the following statements about making important financial decisions, such as a financial investment or taking out a loan? [Strongly agree, Agree, Neither nor, Disagree, Strongly disagree] • I rarely read all the small print before making important financial decisions. • I do a lot of research before making an important financial decision. • I rarely talk to a financial advisor before making an important financial decision. • I definitely talk to family and friends before making an important financial decision.
13	How many minutes does it typically take you to get from your home to your workplace? [Minutes]
13	How many children aged 15 or younger permanently live in your household? [Number] And how many children are aged 10 or younger? [Number] And how many children are aged 5 or younger? [Number]
15	*Each respondent was asked one of the following three questions at the end of the socio-demographics module, if for the last question in that module the preload was not missing, the respondent was assigned to one of the dependent interviewing question versions, and the respondent did not answer that preload was wrong, don't know or refused:* • Which thoughts crossed your mind when thinking about how the number of persons living in your household compared to last year? [Text] • Which thoughts crossed your mind when thinking about how your marital status compared to last year? [Text] • Which thoughts crossed your mind when thinking about how your highest school qualification compared to last year? [Text]
15	*At the end of the labor market activity module all respondents were asked the following question, if their preload was not missing, the respondent was assigned to one of the dependent interviewing question versions, and the respondent did not answer that preload was wrong, don't know or refused:* • Which thoughts crossed your mind when thinking about how your labor market activity compared to last year?

[a] This number matches the question number in Table 4 of the paper