ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Eggs, Johannes; Jäckle, Annette

Working Paper

Dependent interviewing and sub-optimal responding

ISER Working Paper Series, No. 2014-32

Provided in Cooperation with:
Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: Eggs, Johannes; Jäckle, Annette (2014) : Dependent interviewing and sub-optimal responding, ISER Working Paper Series, No. 2014-32, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:
https://hdl.handle.net/10419/126475

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ISER Working Paper Series

Dependent interviewing and sub-optimal responding

Johannes Eggs
Institute for Employment Research
Germany

Annette Jäckle
Institute for Social and Economic Research
University of Essex

No. 2014-32
September 2014

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

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

www.iser.essex.ac.uk

Non-technical summary

The main purpose of panel surveys is to measure changes, or stability, experienced by individuals over time. This is done by asking respondents a core set of questions in the same way at different points in time. Respondents may however respond to a given question differently over time, for various reasons, even if their situation has in fact not changed. As a result panel data tend to over-represent changes experienced by individuals. Proactive dependent interviewing is a questioning method that is used to reduce such spurious changes. The individual answers from the previous interview are preloaded into a computerized questionnaire such that respondents are reminded of their previous response before being asked about their current situation. For example, respondents may be asked "Last time we interviewed you on <date>, you said you were receiving unemployment benefits. Is that still the case?" Reminding the respondent of their previous answer aids their recall and improves the consistency of reports over time. Reminding the respondent of previous answers however also offers opportunities for undesired shortcuts. Respondents may choose to accept the previous information as still applying, regardless of whether it is still an accurate description of their current situation. This could happen if respondents are not sufficiently motivated to make the necessary effort to provide an accurate response or if they have difficulty recalling information and decide that the information presented seems a plausible answer.

In this study we used data from the German Labour Market and Social Security panel study in which an error was made with the preload data about receipt of welfare benefits. The survey data were linked to individual administrative records on receipt of welfare benefits, such that we were able to identify measurement error in the survey reports, and derive measures of the complexity of the respondent's history of receipt that were not affected by measurement error. The results showed that a large proportion of respondents accepted the false preload. This behaviour seemed mainly driven by the difficulty of the response task: respondents with a more complex history of receipt according to the records were more likely to confirm the false preload. Personality also seemed related to the probability of confirming. Predictors of shortcutting behaviour by the respondent that have been found in other studies (including measures of the respondent's cognitive ability, their motivation to respond to the survey, and indicators of shortcutting responses on other items in the survey) and characteristics of the interviewer and the survey were not predictive of confirming the false preload.

Dependent Interviewing and Sub-Optimal Responding

Johannes Eggs
Institute for Employment Research, Germany

Annette Jäckle
ISER, University of Essex

Abstract:

With proactive dependent interviewing respondents are reminded of the answer they gave in the previous interview, before being asked about their current status. We examine the risk that respondents falsely confirm the answers from the previous interview as still applying, using data from a panel survey in which preload data about receipt of welfare benefit contained errors. A large proportion of respondents confirmed the false preload. Respondents with a more complex history of receipt, according to linked administrative records, were more likely to confirm. Personality also seemed to matter. Predictors of satisficing and characteristics of the survey and interviewer were not predictive of confirming the false preload.

Keywords: measurement error, validation, record linkage, panel survey, welfare benefit, satisficing

JEL classification: C81, C83

Corresponding author: Johannes Eggs, Institute for Employment Research, Regensburger Str. 104, 90478 Nürnberg, Germany. Johannes.Eggs@iab.de

Acknowledgements: This work was supported by the Department for Statistical Methods (KEM) at the Institute for Employment Research (IAB), Germany. We thank Mark Trappmann and Tarek Al Baghal for comments on earlier versions of this article.

1. Introduction

With Proactive Dependent Interviewing (PDI), respondents are reminded of the answer to a survey question they gave in a previous interview, before being asked about their current situation (Mathiowetz and McGonagle 2000). For example, "Last time we interviewed you, you told us that you were working as a pharmacist. Is this still the case?" Dependent interviewing questions are implemented by preloading each respondent's answer from the previous interview into the computerized questionnaire script. Variants of dependent interviewing are nowadays used in most longitudinal panel studies (Schoeni et al. 2013). PDI is commonly used to collect information about labour market status and employment characteristics such as industry and occupation (e.g. in the UK Household Longitudinal Study (UKHLS), Current Population Survey (CPS), National Longitudinal Survey of Youth 1997 (NLSY97), Health and Retirement Study (HRS), English Longitudinal Study of Ageing (ELSA), Survey of Labour and Income Dynamics (SLID)). In this paper, we examine the risk that respondents confirm answers from the previous interview, regardless of whether they are accurate or not.

PDI is used for two main reasons (Jäckle 2009). First, PDI questions can be used to determine routing in the questionnaire and to omit redundant questions. For example, if the respondent is still working for the same employer and in the same occupation as at the previous interview, other characteristics of the job may not have to be collected again. Thus, PDI reduces respondent burden, may shorten the interview and facilitates the flow of the interview (Jäckle 2008; Sala, Uhrig and Lynn 2011). Second, PDI increases the longitudinal consistency of responses across interviews. When questions are asked independently, without reference to previous answers, respondents may for various reasons report a different status in one interview from the next, even if their actual status has not changed (Moore et al. 2009). PDI reduces spurious changes in responses over time, by reducing measurement error in each interview (Lynn et al. 2012).

However, the use of PDI can have disadvantages. Concern is voiced that respondents may falsely confirm a previous status as still applying, as they rely on recognizing the previous information instead of retrieving information from memory (Hoogendoorn 2004). Dependent interviewing could thus lead to spurious stability replacing the original problem of spurious change. Also, inaccurate responses from previous interviews may be confirmed by

respondents as still applying, such that errors are carried forward into future interviews (Conrad, Rips and Fricker 2009). Thus, PDI might provide new sources of measurement error, if respondents falsely confirm information from previous interviews.

In this study we use data from the German panel survey "Labour Market and Social Security" (PASS), where preload information regarding welfare receipt was falsely processed for a subgroup of respondents in one panel wave. We use the survey data linked to individual level administrative data on welfare receipt to address the following questions:

(1) To what extent do respondents confirm previous information when that is false? How much of the apparent false confirmation is in fact due to false reporting at the previous wave?
(2) What are the mechanisms causing false confirmation?
(3) Which socio-demographic characteristics are associated with false confirmation?
(4) What are the implications of false confirmation for measurement error?

2. Background

False confirmation, and measurement error in general, is caused by sub-optimal responding (see Thomas 2014). Sub-optimal responding occurs if individuals are not sufficiently motivated to invest the necessary cognitive resources to respond optimally, or if other non-motivational factors related to the question design or survey implementation interfere. Errors can occur in any step of the response process described by Tourangeau, Rips and Rasinski (2000 p. 8): comprehension of the question and response options, retrieval of relevant information from memory, judgment of the retrieved information to form a conclusion, and formulating a response or selecting a response option.

With proactive DI, the respondent has to compare the information they are reminded of with information retrieved from memory and judge both sets of information. The retrieval and judgment of information are burdensome cognitive processes and there are several stages at which sub-optimal responding can occur. Even if respondents are motivated to provide an accurate response, there are several factors that could lead to false confirmation of previous information. Respondents may fail to understand the question or response options. For example, they may be confused about the type of welfare income they are being asked about.

Respondents may have trouble recalling relevant information, which could be because they never encoded the information in memory, or due to memory decay. For example, they may have trouble remembering the names of income sources they have received, or the dates of receipt. In this case they may believe the information from the previous interview to be correct and therefore confirm it. Respondents may further have difficulty judging the retrieved information against the information they are reminded of. For example, they may have difficulty matching the retrieved dates of receipt against the time periods they are being asked about in the interview. Finally, respondents may inadvertently select an inaccurate response option.

If respondents are not sufficiently motivated to provide an accurate response, they may satisfice by choosing a cognitive shortcut in order to minimize effort (Krosnick 1999). There are several satisficing strategies that could lead to false confirmation of previous information. Firstly, respondents may minimize effort by stopping the search for a response at the first plausible endpoint. PDI provides the respondent with an easy plausible endpoint, which is simply confirming the previous information. Alternatively, respondents may be susceptible to a general tendency to agree with information presented to them. According to psychological theory, individuals are more likely to confirm than to reject information or statements. This phenomenon is known as confirmation bias (Nickerson 1998). Similarly, respondents tend to agree to questions out of an inner impulse or in order to be liked or to avoid a conflict or an argument with the authority respectively the interviewer. This response strategy is also known as acquiescence (Johanson and Osborn 2004, p.536; Tourangeau, Rips and Rasinski 2000, p.5).

The likelihood that respondents satisfice by selecting the first plausible response or acquiescing is thought to be higher with respondents who are less motivated to participate in the survey, as they are less likely to be willing to invest the required cognitive resources to provide accurate information (Krosnick 1999). Respondents with lower cognitive abilities have to invest more mental resources to retrieve and formulate an accurate answer and are therefore also more likely to satisfice (Krosnick 1999). More difficult tasks require more cognitive resources and thus an increased difficulty also increases the risk of satisficing (Meisenberg and Williams 2008).

The likelihood of acquiescing is also related to personality (Kieruj and Moors 2013) and survey procedures. Olson and Bilgen (2011) found that more experienced interviewers elicited higher rates of acquiescence than inexperienced interviewers. Survey mode may also influence the degree of acquiescence and sometimes higher levels of acquiescence are found in telephone interviews (de Leeuw 2005). Research indicates that acquiescence is also related to age and sex of the respondent (Vaerenbergh and Thomas 2013). Regarding age and sex of the interviewer and their association with satisficing and acquiescence, research is however inconclusive (Davis et al. 2010).

Cognitive ability and task difficulty may also be related to sub-optimal responding among respondents who are motivated to provide accurate responses (Knäuper et al. 1997). Respondents with higher ability may be more likely to accurately remember information about welfare receipt, and find it easier to accurately compare the retrieved information with the information from the previous interview. Similarly, if the task is more difficult, respondents are more likely to have trouble accurately recalling and judging information.

In sum, sub-optimal responding may lead respondents to confirm information from previous interviews even if it is not correct. This could be due to motivational problems or other factors influencing the response process. Overall, we expect the likelihood that respondents falsely confirm previous information to be higher among respondents who are less motivated to provide accurate information, respondents with lower cognitive ability and if the task set by the survey question is more difficult. In addition, we expect that some personal characteristics and characteristics of the survey may influence the likelihood that respondents falsely confirm previous information.

The extent to which respondents falsely confirm information presented to them in PDI questions is not known. However there is a previous study that examined the confirmation of PDI questions when the preloaded information was wrong. Aughinbaugh and Gardecki (2008) used data from the NLSY97, where the preload information about receipt of a certain type of welfare income was not drawn from the previous wave interview, but from two waves before. For a sub-sample of 610 respondents the receipt status had changed between these two interviews. Thus these respondents were reminded that they had received/not received the welfare income at the date of the previous interview, when in fact they had reported the opposite at the last interview. The authors found that only one third of these respondents

corrected the information presented to them in the PDI question. Respondents with higher scores on an intelligence measure and respondents who were rated as being more honest by the interviewer were more likely to correct the false preload information. A limitation of this study is that the true status of welfare receipt was unknown. For respondents who had misreported their receipt status at the previous wave, the preload information from two waves earlier was in fact correct and respondents would rightly have confirmed the preload. We use the unique opportunity presented by the combination of an error in preload data and linked administrative records, to identify respondents for whom the preload was truly wrong, to examine their reactions to the preload, and to check the implications for measurement error.

Misreporting of welfare receipt is related to the probability of actual receipt, and thus with a range of socio-economic indicators. In a study by Bruckmeier, Müller and Riphahn (2014) that used data from the same survey and linked administrative records as we use in this study, recipients that were more like non-recipients were more likely to under-report receipt, than recipients whose eligibility was certain. For example, respondents where another household member was in work or who had higher levels of household savings were more likely to under-report receipt. Respondent characteristics related to misreporting receipt might also be associated with the risk of falsely confirming. We therefore also examine whether the types of respondents who are more likely to under-report receipt, are also more likely to falsely confirm information presented to them in PDI questions.

3. The panel survey and validation data

The data for this study are from the German panel survey "Labour Market and Social Security" (PASS). The survey was established to study the impact of major welfare reforms, called the “Hartz reforms” that introduced a new type of welfare scheme called unemployment benefit II (UB II). PASS was designed to assess the dynamics of UB II receipt and to investigate how the welfare reforms influence the social situation of affected households and the persons living in them. PASS was set up as a household survey, since UB II provides economic resources that are means tested at the level of the benefit unit. A benefit unit consists of at least one adult plus their spouse (if applicable) plus any dependent children living with them. A benefit unit is in most cases congruent with the household. The panel study is conducted by the Institute for Employment Research and is funded by the German Federal Ministry for Employment and Social Affairs.

3.1 Survey design

In order to compare recipients of UB II with non-recipients, PASS was set up as a dual-frame survey. It consists of a recipient sample and a sample drawn from the general population. The recipient sample was selected from a register of recipients of UB II held by the German Federal Employment Agency. 300 primary sampling units (PSUs) were drawn from postcodes with selection probabilities depending proportionally on the size of the population. Within each PSU, benefit units were drawn. The population sample was based on a commercial database of household addresses, where addresses were sampled within PSUs. The population sample was stratified disproportionately by socio-economic status such that households with low status were oversampled. Subsequently, refreshment samples were drawn every year. The refreshment samples consist of households that are first time recipients of UB II. Sizes of the refreshment samples vary around 1000 households covering around 1400 individuals aged 15 years or older.

Prior to the first survey interview, each household receives an advance letter that informs the household about the study and includes a leaflet describing the data security protocol. To collect information about the household, the head of the household is asked to complete a household interview containing among others questions on household composition and receipt of UB II. For the recipient sample the head of the household is defined as the person that applied for UB II. For the population sample, the head of the household is defined as the person that is most familiar with the overall situation of the household. After the household interview, every member of the household aged fifteen or older is asked to complete a personal interview. Proxy interviews for currently unavailable members of the household are not allowed.

PASS uses a mixed mode design whereby data are collected using either computer-assisted telephone interviews (CATI) or computer-assisted personal interviews (CAPI). In wave 1 households were first approached in CATI, non-respondents and households for whom no valid telephone numbers were known were followed up with CAPI. From wave 2 onwards households are first approached in the mode in which they were last interviewed. Refreshment samples are contacted first by CAPI. The first time a household is interviewed, each household member who completes the personal interview receives a conditional incentive of 10 Euros. In subsequent panel waves, the incentive is posted unconditionally

together with the advance letter that informs respondents of the upcoming interview. In order to assess socio-economic dynamics, households are interviewed annually. In wave 1 PASS had household response rates of 28.7% for the recipient sample and 24.7% for the population sample (RR1 according to The American Association for Public Opinion Research 2011). For an overview of the PASS panel see Trappmann et al. (2013).

3.2 Administrative data and linkage

The administrative data used to validate survey reports are from the Integrated Employment Biographies (IEB) held by the Research Data Centre of the German Federal Employment Agency. It contains exact start and end dates of all spells of UB II receipt. This information is of high quality as it is directly produced by the software that administers benefit claims and payments (Jacobebbinghaus and Seth 2007; Köhler and Thomsen 2009). The IEB is a person level dataset. Spells that refer to a benefit unit are therefore recorded for each person in that unit.

The linkage between PASS survey data and IEB administrative data requires informed consent of respondents. Respondents who have not given consent to data linkage are asked again in the following wave. Among respondents interviewed at wave 4 (the wave we focus on in this study), 81% had given consent to linkage at some point. The recipient sample was selected from the IEB data and therefore linkage was trivial. Respondents in the population sample were linked by their name and address, gender and date of birth using error tolerant procedures based on Jaro (1989).

3.3 Dependent interviewing and preload error

PASS uses PDI to collect information on UB II receipt. As UB II is a means tested welfare programme that is paid to households, the information is collected in the household questionnaire. The head of the household is asked:

In the last interview in <MONTH/YEAR> you stated that the household you were living in then was receiving unemployment benefit 2 at the time. Until when was this benefit received without interruption? Please report the month and the year.

PDI depends on preload information. For this question, the preload is whether or not the household was receiving UB II at the time of the previous interview. When preparing the

preload information for wave 4 an error occurred: households that reported a terminated UB II spell, but no current receipt at the wave 3 interview, were coded as still receiving UB II at the time of the interview. In the PDI question these households were reminded that they had received UB II at the time of the interview and asked until when it had continued, when in fact they had reported that receipt had ended by then. If the respondent said that the preload information was wrong, the spell was treated as having ended at the previous interview date and the respondent was asked whether they had had any other spells of receipt since. That is, respondents were not explicitly asked to confirm the preload, but if they disputed the preload data this was treated as a valid response. We use the expression "confirmed the preload" somewhat loosely to refer to respondents who did not contradict the preload. The preload error occurred for 393 households; 73.7% from the recipient sample, 11.1% from the population sample, 15.2% from the refreshment samples. These households form the base for our analyses and 354 were successfully linked to administrative data.

4. Predictors of sub-optimal responding

Sub-optimal responding is related to the cognitive ability of the respondent, the difficulty of the response task, and the motivation of the respondent to provide accurate information.

As proxy measures for *cognitive ability* we use education and age. A dichotomous indicator is set to one, if the respondent holds an intermediate or higher degree. We expect that having a degree is negatively associated with confirming the false preload. Cognitive ability decreases with age. However as our study sample consists of individuals below 67 (the age cut-off for UB II eligibility), we expect only a weak association between age and confirming the false preload.

How *difficult* the task of reporting on UB II receipt is for the respondent depends on the complexity of their history of receipt. Respondents who have had multiple spells of receipt will find it more difficult to accurately recall details of any one particular spell (Eisenhower, Mathiowetz and Morganstein 1991). The administrative records of 354 households could be used to derive two indicators of the complexity of the respondent's history: the number of spells of UB II receipt and the elapsed time since receipt ended. We use the number of welfare spells for the time period of 12 months around the date of the wave 3 interview. We expect that the number is positively related to the confirmation of the false preload, as the increased complexity of the respondent's history makes it more likely that the respondent will

make errors in recalling information or that they will not expend the necessary cognitive resources to accurately assess the possibility of welfare receipt at the time of the last interview. The elapsed time measures the time between the end of the last spell of UB II receipt and the date of the last interview. We expect that the elapsed time is negatively associated with the false confirmation. If more time has passed between welfare receipt and interview date, it should be easier for the respondent to remember correctly, whether welfare was received at the time of the last interview. Thus respondents should be less likely to make errors of recall and judgment and should need less effort to report accurately, reducing the probability of sub-optimal responding.

We further use interviewer observations as proxies for the combined effect of respondent cognitive ability and difficulty of the response task (questions in the Appendix). Interviewers were asked on a 5-point scale, whether the respondent had difficulty remembering dates. The variable was coded as 1 if the interviewer judged that the respondent had difficulty or strong difficulty remembering dates. We expect the interviewer judgment of whether respondents had difficulty recalling information to be positively associated with confirming the false preload, as respondent difficulty could be due to low cognitive ability or a complex history of receipt, or both, which would increase the likelihood of sub-optimal responding. The indicators of ability (education, age) and complexity of the respondent's history (number of spells in the records, elapsed time since end of receipt) are correlated to some extent with the interviewer assessments of whether the respondent had difficulty recalling dates of events. The largest correlation is between difficulty dating events and education (-0.14, p=0.02), suggesting that the interviewer observations do measure additional aspects related to ability and difficulty.

The *motivation* of respondents is measured by observations made by the interviewer, as well as indicators of satisficing on other items in the survey (Hoogendoorn 2004). Interviewers were asked on a 5-point scale whether they believed that the respondent was interested in the interview. The variable was coded as 1, if the respondent had shown no or little interest. A similar strategy was chosen by Aughinbaugh and Gardecki (2008). We further use the amount of rounding, non-differentiation and "don't know/refused" answers by the respondent, which are commonly used proxies for satisficing (Krosnick et al. 2002). Dichotomous indicators are formed that were coded as 1 if the respondent rounded in more than 50% of the questions in the household questionnaire asking about monetary amounts

(where a responses was classified as rounded if it was a multiple of 50 Euros), used constantly the same response option in at least one of three item batteries, respectively had more than 1% “don’t know/refused” answers in the personal questionnaire. On average each respondent received six numerical questions and 103 questions in the survey. We expect low motivation and the indicators of satisficing on other items to be positively associated with confirming the false preload.

Additional indicators related to *acquiescence* include personality traits and characteristics of the survey and interviewers. Acquiescence is related to agreeableness (Knowles and Nathan 1997). Agreeableness is one dimension of the “Big Five” personality traits. The Big Five are broad dimensions that depict the range of personalities (John and Srivastava 1999). The personality traits are measured by a German version of the Big Five item battery (Rammstedt and John 2005). These dimensions are the traits of extroversion, agreeableness, conscientiousness, neuroticism, and openness. Factor scores were calculated for each dimension via confirmatory factor analysis in line with Rammstedt and John (2005). We focus on agreeableness and expect that a higher agreeableness score is positively related with confirming the false preload. The Big Five item battery was only measured one wave after the preload error occurred. It has however been argued that acquiescence is a stable personality trait (Kieruj and Moors 2013). Hence, the later data collection should not distort the hypothesized relation between the constructs. However, cases are lost due to panel attrition from wave 4 to wave 5. Survey and interviewer-specific characteristics can also influence acquiescence. We expect telephone interviewing (versus face-to-face) and interviewer experience to be positively associated with confirming the false preload.

Finally, previous research has shown that the risk of measurement error in reporting welfare receipt is associated with socio-economic factors (Bollinger and David 1997; Bruckmeier, Müller and Riphahn 2014). Bruckmeier et al showed that women, singles, younger individuals, individuals in higher income categories, with larger amounts of savings and shorter spells of welfare receipt were more likely to misreport. They concluded that respondents that were less likely to receive welfare were more likely to under-report. The authors also used data from the PASS panel survey. Hence, we derived similar indicators as in this earlier study to test whether the indicators related to under-reporting are also associated with the risk of confirming the false preload.

5. Results

To what extent do respondents confirm previous information when that is false?

For our analyses we use the interviews of 393 heads of households, who at wave 4 received a question with false preload information regarding their welfare receipt at the time of the wave 3 interview. Of these, 30.1% contradicted the interviewer, stating that the preloaded information was false. That is, 69.9% of respondents did not correct the preload. Instead they either reported that the spell had ended between the wave 3 interview and the wave 4 interview (46.8%), or was still ongoing at date of the wave 4 interview (17.8%), or that the spell had ended and a new one had started (5.3%).

How much of the apparent false confirmation is due to false reporting at the previous wave?

All respondents included in our analysis sample reported at the previous interview that they were not currently receiving UB II. However, some of these respondents may have under-reported receipt. Welfare receipt can be considered a sensitive item that is generally under-reported in social surveys (Bound, Brown and Mathiowetz 2001). In the PASS survey, UB II is under-reported by about 10-15% (Kreuter et al. 2010). Therefore for some respondents in our analysis sample, the apparently false preload indicating receipt at the time of the previous interview may in fact have been correct and these households would have been correct in confirming the preload. We can identify households that under-reported welfare receipt at the previous interview using the register data. Table 1 documents the extent to which respondents confirmed the preload, by whether the preload was in fact correct. Of the 354 households that could be linked, 74 (20.9%) had according to the records received UB II at the time of the last interview. Of these households 68 (91.9%) confirmed the preload. In contrast, among the 280 households where the preload really was wrong, only 64.3% confirmed the preload. The probability of confirming the preload was therefore significantly higher if the preload was in fact correct ($P<0.001$). However of the overall confirmation rate of 70.1%, only 19.2 percentage points were due to respondents who under-reported receipt at the previous wave. The remaining 50.9% were respondents who confirmed a preload that really was wrong.

Table 1: Probability of confirming preload, by whether preload was correct

	Confirmed preload:					
	Yes		No		Total	
Validation against records:	n	row %	n	row %	n	col %
Preload wrong	180	64.3	100	35.7	280	79.1
Preload correct	68	91.9	6	8.1	74	20.9
Total	248	70.1	106	29.9	354	100.0

Notes: Chi(2)=21.84, P<0.001

What are the mechanisms causing false confirmation?

For the subsequent analyses we focus on the 280 households where the preload really was wrong according to the records, and respondents were reminded of receipt when in fact they had not been receiving UB II at the date of the previous interview. The descriptive statistics for this subgroup are shown in appendix tables A1 and A2.

To test which mechanisms might explain why respondents confirm false preload information, we first test the bivariate associations between each of the predictors of sub-optimal responding and the probability of confirming the preload (Tables 2 and 3). We split continuous variables at the mean or into quintiles and use $\chi 2$ tests to test for significant associations. We then estimate multilevel logistic models and calculate average marginal effects for the probability of confirming the false preload (Table 4). The 280 respondents are nested in 170 interviewers; 79 of the interviewers conducted only one interview with a respondent from the analysis sample, while 91 interviewers conducted two or more interviews. We include the interviewer level to estimate standard errors of interviewer level variables appropriately; we do however not interpret interviewer effects, due to the small number of respondents per interviewer. As the Big Five personality traits were collected a year after the preload error, and hence some observations are lost to attrition, we estimate separate models excluding and including the Big Five traits.

Our measures of respondent *cognitive ability* were not significant predictors of the probability of confirming the false preload. While there was a tendency for respondents with lower education to be more likely to confirm the preload than respondents with higher education, this difference was not significant in the bivariate (Table 2) or multivariate (Table 4) tests. Similarly, while there were some differences between age quintiles in the probability of

confirming (Table 3), there was no clear pattern in the effects and the probability of confirming did not appear to increase with age as expected.

Table 2: Bivariate associations, confirming false preload and predictors of sub-optimal responding

	Percent confirmed preload	Test of proportions (*p*-value)	n
Higher education = 0	68.9	0.122	119
Higher education = 1	59.9		157
Respondent age ≤ 55	63.8	0.996	218
Respondent age > 55	63.8		58
Number of UB II spells in records ≤ 2	60.4	0.002	235
Number of UB II spells in records > 2	84.4		45
Months since last UB II receipt in records ≤ 6.8	76.1	0.000	117
Months since last UB II receipt in records > 6.8	55.8		163
Difficulty dating events = 0	62.7	0.322	244
Difficulty dating events = 1	75.0		16
Interview not interesting = 0	61.5	0.591	122
Interview not interesting = 1	64.7		136
Rounding in more than 50% of questions = 0	66.3	0.108	211
Rounding in more than 50% of questions = 1	55.4		65
Non-differentiation in 1+ item batteries = 0	64.8	0.627	179
Non-differentiation in 1+ item batteries = 1	61.9		97
Item non-response ≤ 1%	63.2	0.605	239
Item non-response > 1%	67.6		37
CATI	61.9	0.202	194
CAPI	69.8		86
Male interviewer	72.0	0.021	118
Female interviewer	58.6		162
Interviewer experience ≤ 3 months	68.1	0.122	160
Interviewer experience > 3 months	59.2		120
Agreeableness score ≤0	64.2	0.675	109
Agreeableness score > 0	67.0		94
Extraversion score ≤ 0	65.0	0.879	100
Extraversion score > 0	66.0		103
Openness score ≤ 0	70.1	0.313	77
Openness score > 0	63.2		125
Neuroticism score ≤ 0	60.5	0.090	114
Neuroticism score > 0	71.9		89
Conscientiousness score ≤ 0	66.7	0.744	96
Conscientiousness score > 0	64.5		107

Notes: Continuous variables split at the mean.

Table 3: Bivariate associations, confirming false preload and predictors of sub-optimal responding (continuous variables by quintiles)

	Percent confirmed preload	n	χ^2 test (*p*-Value)
Respondent age 20 - 32	64.3	56	
Respondent age 33 – 40	53.6	56	
Respondent age 41 – 48	60.3	63	
Respondent age 49 – 56	80.0	50	
Respondent age 57 - 67	62.8	51	0.073
1 - 3 months since last UB II receipt in records	80.4	56	
4 - 6 months since last UB II receipt in records	75.0	56	
7 - 8 months since last UB II receipt in records	60.7	56	
9 – 13 months since last UB II receipt in records	57.1	56	
14 - 39 months since last UB II receipt in records	48.2	56	0.002
Interviewer experience 1 - 2 years	64.9	94	
Interviewer experience 3 years	72.7	66	
Interviewer experience 4 years	70.0	40	
Interviewer experience 5 years	51.4	35	
Interviewer experience 6+ years	55.6	45	0.156

The measures of task difficulty derived from the administrative records were strong predictors of the probability of confirming the false preload. Respondents with two or more spells of UB II receipt in the 12-month window around the wave 3 interview were 24 percentage points more likely to confirm the false preload according to the bivariate test (Table 2, p=0.002), than respondents with one or no spell. Controlling for other characteristics, the average marginal effect estimated from the multivariate model (Model 1 in Table 4) suggests that each additional spell of UB II receipt increased the probability of confirming the false preload by 18.1% (p<0.001). Similarly, respondents for whom the length of time between the end of the last UB II spell and the date of the wave 3 interview was shorter than the average of 6.8 months, were 20.3 percentage points more likely to confirm the false preload according to the bivariate tests (Table 2, p=0.000) than respondents whose elapsed time was longer than average. Examining the probability of confirming the preload by quintiles of the elapsed time shows a clear linear relationship (Table 3): the probability of confirming was highest amongst those where the elapsed time was only 1 to 3 months (80.4%), and monotonically fell to 48.2% among the group with the longest elapsed time of 14 - 39 months (p=0.002). These results are confirmed by the multivariate estimates (model 1 in Table 4) according to which each additional month between the end of the spell and the date of interview decreased the probability of confirming the preload by 1.1% (p<0.05).

Table 4: Average marginal effects of random effects logistic models for confirming false preload

	(1)		(2)		(3)	
Pr(confirmed false preload)	AME	se	AME	se	AME	se
Higher education	-0.126	0.077	-0.079	0.103	-0.087	0.109
Respondent age	0.004	0.003	0.004	0.004	0.004	0.004
Number of UB II spells in records	0.181***	0.043	0.155**	0.056	0.155**	0.056
Months since last UB II receipt in records	-0.011*	0.005	-0.015*	0.007	-0.015*	0.007
Interviewer: difficulty dating events	0.207	0.159	0.551	0.297	0.595	0.313
Interviewer: interview not interesting	0.049	0.074	0.023	0.090	0.010	0.092
Rounding in > 50% of questions	-0.075	0.085	-0.039	0.113	-0.025	0.112
Non-differentiation in 1+ item batteries	0.030	0.104	-0.054	0.144	-0.093	0.149
Item non-response > 1%	-0.008	0.077	-0.003	0.107	0.020	0.109
CAPI (vs. CATI)	0.113	0.097	0.010	0.141	0.028	0.137
Female interviewer	-0.101	0.088	-0.198	0.123	-0.190	0.119
Interviewer experience in years	0.004	0.020	0.015	0.027	0.013	0.026
Agreeableness			0.139*	0.068	0.108	0.067
Extroversion					0.015	0.058
Openness					-0.019	0.061
Neuroticism					0.094	0.061
Conscientiousness					-0.052	0.078
Rho	0.39		0.56		0.57	
Observations	242		177		176	
AIC	353.4		265.3		281.4	

Notes: * p<0.05, ** p<0.01, *** p<0.001

The interviewer assessment of whether the respondent had difficulty recalling dates of events was not significantly associated with the probability of confirming the preload. While respondents who were judged to have had difficulty tended to be more likely to confirm the preload (Table 2) the difference was not significant and not confirmed by the multivariate estimates.

The indicators of respondent motivation were also not associated with the probability of confirming the preload. According to the bivariate tests (Table 2) and the multivariate estimates (Table 4) there were no differences in the probability of confirming regardless of whether or not the interviewer judged that the respondent had shown little interest in the survey, and whether or not the respondent had rounded, non-differentiated or given don't know or refusal responses to other items in the questionnaire.

Characteristics of the interviewer and survey that may be related to acquiescence were also not associated with the probability of confirming the false preload. Although respondents interviewed by men were 13.4 percentage points more likely to confirm the preload than respondents interviewed by women in the bivariate tests (Table 2, p=0.021), interviewer sex was not significant in the multivariate model (Table 4). Interviewer experience was not related to the probability of confirming the preload in the bivariate tests (whether split at the mean in Table 2, or split by quintile in Table 3) or the multivariate test. The survey mode was also not significantly associated with the probability of confirming in any of the tests.

Respondent personality was associated with the probability of confirming the false preload. In the bivariate tests, where the indicators for personality traits were dichotomized at the mean, agreeableness was not associated with the probability of confirming. However controlling for other characteristics and including the agreeableness score as a continuous variable in the multivariate test (Table 4, Model 2), each additional point on the agreeableness score (which ranged from -1.70 to 1.61) increased the probability of confirming the false preload by 13.9% (p<0.05). However when the other four personality traits, extroversion, openness, neuroticism and conscientiousness, were included in the model (Model 3), none of the traits were significant predictors of confirming the preload. As the Big Five measures were collected in a subsequent wave and cases were lost due to attrition, we estimated additional models using the independent variables from model 1 and the estimation samples from models 2 and 3 to check for selectivity in the results due to attrition. There were no relevant shifts in the results (not shown) and thus we assume that the results are robust to the sample selection due to attrition. Using the estimation sample for model 3 to rerun models 1 and 2 further suggests that model fit did not improve much by adding personality traits: the Akaike Information Criterion (AIC) changed from 263.0 in Model 1 to 264.0 in Model 2 when agreeableness was added, and to 281.4 when the remaining Big Five traits were added. This suggests that while personality may have had some effect leading to acquiescence, the complexity of the respondent's history was the main driver of confirming the false preload.

In sum, while a large proportion of respondents confirmed the false preload, this behaviour does not seem to be driven by lack of respondent motivation. Indicators of motivation and satisficing on other items, and indicators of respondent cognitive ability were not predictive of confirming the preload. Instead, the measures of the difficulty of the response task derived

from administrative records were strong predictors: those respondents for whom the task of recalling information about any one particular spell would have been more difficult were more likely to confirm the false preload. In addition, respondents who scored higher on agreeableness were more likely to confirm the false preload.

Who is at risk of falsely confirming previous information?

Previous research has shown that specific socio-economic groups are more likely to misreport their welfare receipt status. Using the indicators that predicted under-reporting of UB II receipt in the study by Bruckmeier, Müller and Riphahn (2014), we tested whether the same factors increased the risk of confirming the false preload. The predictors included the respondent's sex, age, whether they had a disability, whether they were an immigrant, education, household type, whether anyone in the household was in regular employment, monthly net household income, value of household savings, whether they owned their home, the number of months of receipt of UB II over the life of the panel, and location (East or West Germany). We estimated multilevel logistic regression models for the probability of confirming the false preload. Respondents were nested in interviewers. The results are presented in table A3 in the Appendix. The results show no significant associations between the socio-economic indicators and confirming the false preload. The confirmation of false preloads therefore seems to be driven by different factors than under-reporting of receipt.

What are the implications of the respondent behaviour for measurement error?

The false preload reminded respondents that they had received UB II at the time of the previous interview, although at the time they had reported that receipt had ended. For respondents who confirmed this false information, the error in receipt status may therefore be carried over to the current interview. We therefore also examine what impact confirming the false information from wave 3 had on measurement error in receipt status at the time of the wave 4 interview (Table 5). Among all respondents for whom the preload error was made, the wave 4 receipt status was wrong for 11.2%. As expected, the error rate was higher for respondents who confirmed the preload (14.4%), than for those who did not confirm (3.8%). We would expect the error due to confirmation of the false preload to mainly consist of over-reporting: respondents who confirmed the false information that they were receiving UB II at the time of the wave 3 interview would be likely to over-report receipt at the wave 4 interview. Surprisingly however, while 29 of the misreporters over-reported receipt, 10

under-reported. That is, these respondents reported that the wave 3 spell had ended and failed to report a subsequent spell that, according to the records, was ongoing at the time of the wave 4 interview. A second surprising result is that the error rates were lower when the respondents who had misreported their wave 3 status, such that the preload was actually correct, were excluded. Excluding these cases the wave 4 status was wrong for 8.0% of respondents, with all but one being respondents who had confirmed the preload. This suggests that respondents who misreported at wave 3 were likely to again misreport at wave 4. In sum, respondents who confirmed the false preload were more likely to report their wave 4 status with error, than respondents who did not confirm the preload.

Table 5: Impact of confirming preload on measurement error

		Error in wave 4 status	
	Preload confirmed	Percent	n
Preload error (N=354)	Yes	14.4	243
	No	3.8	106
	Total	11.2 ~	349
Preload wrong according to records (N=280)	Yes	12.0	175
	No	1.0	100
	Total	8.0 †	275

Notes: 5 households missing due to don't know/refusal answer about current receipt.
~ 10 under-reporters and 29 over-reporters; † 9 under-reporters and 13 over-reporters.

6. Discussion

One of the main concerns against using proactive dependent interviewing is that reminding respondents of an answer they gave in a previous interview, before asking about their current status, offers respondents the opportunity to satisfice: respondents might say that the previous answer still applies, regardless of whether their situation has in fact changed. If respondents falsely confirm previous information as still applying, PDI may lead to under-reporting of change.

In this study we present novel evidence on the risk that respondents confirm false information from previous interviews. We use a unique data source combining responses from a panel survey, where the preload data for a PDI question contained errors, with linked individual-level administrative records. We exploit the linked administrative records to identify

measurement error in the survey reports, and to derive indicators not affected by measurement error that describe the respondent's history. Using the combined data we examine the extent to which respondents confirm the false preload, which mechanisms lead respondents to confirm, and the implications for measurement error.

While a large proportion of respondents confirmed the false preload, this behaviour seems mainly driven by recall difficulties among respondents with complex histories, rather than by satisficing behaviours. Overall, 69.9% of respondents confirmed the preload. However using the linked administrative data we were able to identify that the preload, that mistakenly reminded respondents of UB II receipt at the time of the previous interview, was in fact correct for some respondents who had under-reported receipt at the previous wave. Respondents for whom the preload was in fact correct were more likely to confirm the preload, than respondents for whom the preload really was wrong. Nonetheless, the confirmation rate among respondents where the preload really was wrong was still high at 64.3%.

To examine the mechanisms that lead respondents to confirm the false preload, we tested a range of factors that are related to sub-optimal responding. Our results suggest that the confirmation bias was not related to respondent motivation or ability: the probability of confirming the false preload was not related to interviewer observations of respondent interest in the survey, indicators of satisficing on other items in the questionnaire, age, education, or interviewer observations about whether the respondent had recall difficulties. The probability of confirming was also not associated with characteristics of the survey and interviewer (sex, experience and mode of interview) that were related to acquiescence in other studies. Instead the complexity of the respondent's history of welfare receipt was a strong predictor of confirming the false preload. Respondents who, according to the administrative records, had had a larger number of spells of receipt, or for whom the spell had ended close to the date of the previous interview, were more likely to confirm the false preload. This suggests that respondents who would have had difficulty recalling information about any one particular spell were more likely to think that the preload information was plausible and therefore confirm it. The respondent's personality also appeared to have an effect: respondents who scored higher on the agreeableness score were more likely to confirm. However the effect disappeared once other personality traits were controlled for.

We found no associations between the probability of confirming the false preload and socio-economic characteristics that are commonly associated with under-reporting welfare receipt. This suggests that under-reporting and confirmation of false preload information are driven by different mechanisms: respondents who are similar to non-recipients in their socio-economic characteristics are more likely to under-report receipt (Bruckmeier, Müller and Riphahn 2014). This is akin to the common result that those who over-report voting tend to have characteristics similar to voters (Ansolabehere and Hersh 2012). Confirming false preload information however seems to be driven by the complexity of the respondent's history that makes it difficult to report accurately. That is, confirmation is not driven by factors related to group identity or membership, but by the respondent's actual experiences. Testing for links between respondent experiences and reporting errors requires exogenous information about experiences that are not themselves affected by reporting error. We were fortunate to have access to the administrative records as an exogenous source of information about respondents' histories.

We also used the administrative records to examine the implications of confirming the false preload for measurement error. While a majority confirmed the false preload, the current receipt status was wrong for only 11.2% of respondents who had confirmed. The error rate was higher among respondents who had also misreported their status at the previous interview.

Our study has several limitations. First, the results are specific to a sample who had recently reported welfare receipt. Although the preload information was wrong, it was plausible for these respondents, which may explain the high rates of confirming. In Aughinbaugh and Gardecki's (2008) study, where preload errors were also made for respondents who had reported receipt in either of the previous two interviews, the confirmation rates were similarly high. For non-recipients a false preload indicating receipt would be implausible and they would be less likely to confirm it as a response. Confirmation rates are therefore likely to be much lower in general population samples. Investigating the risks of false confirmation in a general population sample would ideally require an experimental design allocating randomized preloads to respondents, where the responses and preloads can be linked to administrative records. Second, the results are specific to the German welfare programme UB II, to the question wording and the reference period. Nonetheless Aughinbaugh and Gardecki (2008) reported similar confirmation rates for a different outcome and with different question

wording and reference period. Third, individuals that agreed to the record linkage are a selective subsample (Beste 2011). In the PASS study respondents that are older and report a higher income are more likely to consent. Fourth, the interviewers' assessments of the respondents' motivation and cognitive difficulties were not associated with the probability of confirming. This could in part be due to measurement problems with the interviewer observations. Previous studies have found mixed results as to the usefulness of interviewer observations. For example, Feldman, Hyman and Hart (1951) found little reliability in interviewer assessments of respondents' intelligence. However, Aughinbaugh and Gardecki (2008) found that respondents rated as being more honest were more likely to correct the false preload and Barret, Sloan and Wright (2006) found that interviewer assessments of the respondent's cognition was positively related to several indicators of data quality. Fifth, there are sizable intra-interviewer correlations in the probability of confirming the false preload. However, as the maximum number of interviews per interviewer is seven and a large number of interviewers conducted only one interview in the analysis sample, a meaningful interpretation of the intra-interviewer correlation is not feasible (Hox 2010). The interviewer effects might also be confounded with area effects for CAPI, although only 41 % of all cases were interviewed via CAPI.

In sum, our study suggests that respondents do not react to the information presented to them in PDI questions by satisficing. The gains achieved by PDI in reducing under-reporting are likely to outweigh the potential costs of false confirmation. This corresponds to conclusions drawn by Lynn et al. (2012) who experimentally contrasted dependent interviewing with independent questions on benefit receipt, where the responses were also linked to administrative records. Their results showed that PDI substantially reduced under-reporting, but did not increase over-reporting of receipt. Our results nonetheless reinforce the need for strict quality control of preload answers used for dependent interviewing questions.

References

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20:437-79.

Aughinbaugh, Alison, and Rosella M. Gardecki 2008. "Assessing Acquiescence among Participants in the National Longitudinal Survey of Youth 1997."

Barret, Kristen, Matt Sloan, and Debra Wright. 2006. "Interviewer Perceptions of Interview Quality." *Proceedings of the ASA - Survey Methods Section*.

Beste, Jonas. 2011. "Selektivitätsprozesse bei der Verknüpfung von Befragungs- mit Prozessdaten. Record Linkage mit Daten des Panels "Arbeitsmarkt und soziale Sicherung" und administrativen Daten der Bundesagentur für Arbeit." in *FDZ-Methodenreport*.

Bollinger, Christopher R., and Martin H. David. 1997. "Modelling discrete choice with response error: food stamp participation." *Journal of the American Statistical Association* 92(439):827-35.

Bound, John, Charles Brown, and Nancy A. Mathiowetz. 2001. "Measurement Error In Survey Data." in *Handbook of Econometrics*, edited by J. J. Heckman and E. Leamer: Elsevier.

Bruckmeier, Kerstin, Gerrit Müller, and Regina T. Riphahn. 2014. "Who misreports welfare receipt in surveys?" *Applied Economics Letters* 21(12):812-16.

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

Davis, R. E., M. P. Couper, N. K. Janz, C. H. Caldwell, and K Resnicow. 2010. "Interviewer effects in public health surveys." *Health Education Research* 25(1):14-26.

de Leeuw, Edith. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21(2):233-55.

Eisenhower, Donna, Nancy A. Mathiowetz, and David Morganstein. 1991. "Recall Error: Sources and Bias Reduction Techniques." Pp. 127-44 in *Measurement Errors in Surveys*, edited by Paul B. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. New York: Wiley.

Feldman, John J., Herbert Hyman, and Carl W. Hart. 1951. "A Field Study of Interviewer Effects on the Quality of Survey Data." *Public Opinion Quarterly* 15(4):734-61.

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.

Hox, Joop J. 2010. *Multilevel Analysis - Techniques and Applications*. New York: Routledge.

Jäckle, Annette. 2008. "Dependent Interviewing: Effects on Respondent Burden and Efficiency of Data Collection." *Journal of Official Statistics* 24(3):411-30.

—. 2009. "Dependent Interviewing: A Framework and Application to Current Research." in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. Chichester: Wiley.

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.

Jaro, Matthew A. 1989. "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida." *Journal of the American Statistical Association* 84(406):414-20.

Johanson, George A., and Cynthia J. Osborn. 2004. "Acquiescence as differential person functioning." *Assessment & Evaluation in Higher Education* 29(5):535-48.

John, Oliver P., and Sanjay Srivastava. 1999. "The Big Five trait taxonomy: History, measurement, and theoretical perspectives." Pp. 102-38 in *Handbook of personality: Theory and research*, edited by L. A. Pervin and O. P. John. New York: Guilford.

Kieruj, Natalia D., and Guy Moors. 2013. "Response style behavior: Question format dependent or personal style?" *Quality & Quantity* 47(1):193-211.

Knäuper, Bärbel, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog. 1997. "Question Difficulty and Respondents' Cognitive ability: The effect on data quality." *Journal of Official Statistics* 13(2):181-99.

Knowles, Eric S., and Kobi T. Nathan. 1997. "Acquiescent responding in Self-Reports: Cognitive Style or Social Concern." *Journal of Research in Personality* 31:293-301.

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.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50(1):537-67.

Krosnick, Jon A. , Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, Stanley Presser, Paula Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. 2002. "The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66(3):371-403.

Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2012. "The impact of questioning 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.

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

Meisenberg, Gerhard, and Amandy Williams. 2008. "Are acquiescent and extreme response styles related to low intelligence and education?" *Personality and Individual Differences* 44(7):1539–50.

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

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2(2):175-220.

Olson, Kristen , and Ipek Bilgen. 2011. "The Role Of Interviewer Experience on Acquiescence." *Public Opinion Quarterly* 75(1):99–114.

Rammstedt, Beatrice, and Oliver P. John. 2005. "Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit." *Diagnostica* 51(4):195-206.

Sala, Emanuela , S.C. Noah Uhrig, and Peter Lynn. 2011. "‘‘It Is Time Computers Do Clever Things!’’: The Impact of Dependent Interviewing on Interviewer Burden." *Field Methods* 23(1):3-23.

Schoeni, Robert F., Frank Stafford, Katherine A. McGonagle, and Patricia Andreski. 2013. "Response Rates in National Panel Surveys." *The ANNALS of the American Acadamy of Political and Social Science* 645:60-87.

The American Association for Public Opinion Research. 2011. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition." AAPOR.

Thomas, Randall K. 2014. "Fast and Furious … … or Much Ado About Nothing? Sub-Optimal Respondent Behavior and Data Quality." *Journal of Advertising Research* 54(1):17-31.

Tourangeau, Roger, Lance Rips, J., and Kenneth Rasinski. 2000. *The Psychologie of Survey Response*. Cambridge: Cambridge University Press.

Trappmann, Mark , Jonas Beste, Arne Bethmann, and Gerrit Müller. 2013. "The PASS panel survey after six waves." *Journal for Labour Market Research* 46(4):275-81.

Vaerenbergh, Yves Van, and Troy D. Thomas. 2013. "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public Opinion Research* 25(2):195-217.

Appendix

Interviewer observations, asked at the end of each personal interview:

In your opinion: How difficult was it for the respondent to date certain events?
1 Very difficult
2
3
4
5 Not difficult at all

In your opinion: How interesting was the interview for the respondent?
1 Not interesting at all
2
3
4
5 Very interesting

Table A1: Summary statistics for respondents with false preload (continuous variables)

	mean	sd	min	max	count
Respondent age	44.21	12.23	20.00	67.00	276
Number of UB II spells in records	1.63	0.94	0.00	5.00	280
Months since last UB II receipt in records	9.11	6.41	0.07	39.63	280
Extroversion	-0.05	0.89	-2.20	1.45	203
Openness	0.15	0.81	-2.12	1.40	202
Neuroticism	-0.03	0.81	-1.56	2.10	203
Conscientiousness	0.06	0.70	-2.09	1.12	203
Agreeableness	-0.08	0.72	-1.70	1.61	202
Interviewer experience in years	3.64	2.31	1.00	19.00	280

Notes: The Big Five personality traits were collected in wave 5 and hence some observations were lost due to attrition.

Table A2: Summary statistics for respondents with false preload (categorical variables)

	Percent	count
Female respondent	55.1	276
Higher education	56.9	276
Interviewer observation: difficulty dating events	6.2	260
Interviewer observation: interview not interesting	52.7	258
Rounding in more than 50% of questions	23.6	276
Non-differentiation in 1+ item batteries	35.1	276
Item non-response > 1%	13.4	276
CAPI (vs. CATI)	30.7	280
Female interviewer	57.9	280

Table A3: Average marginal effects of random effects logistic models for confirming the preload (Socio-economic characteristics)

Pr(confirmed false preload)	AME	se
Female respondent	-0.023	0.074
Respondent age	0.001	0.003
Disability	0.056	0.095
Migrated	-0.020	0.120
Omitted: no schooling		
Lower secondary degree	-0.074	0.164
Higher secondary degree	-0.095	0.165
Vocational education	-0.050	0.096
Young children in household (age<=4)	-0.177	0.135
Omitted: single person		
Household without children	0.043	0.114
Single Parent	0.104	0.114
Household with children	0.098	0.107
Other	0.051	0.223
Regular employed person in HH	-0.128	0.082
Omitted: HH income < 500 Euro		
HH income 500 – 749 Euro	0.011	0.223
HH income 750 – 999 Euro	0.113	0.238
HH income ≥ 1000 Euro	0.082	0.226
Omitted: no HH savings		
HH savings < 1000 Euro	0.023	0.081
HH savings 1000 – 2499 Euro	-0.043	0.110
HH savings 2500 – 4999 Euro	0.143	0.147
HH savings ≥ 5000 Euro	-0.107	0.115
HH owns home	-0.153	0.099
Omitted: UB II receipt < 12 months		
UB II receipt 12 – 25 months	-0.081	0.157
UB II receipt > 25 months	0.167	0.137
Eastern Germany	0.094	0.073
N	262	
Rho	0.542	
AIC	353.775	

Notes: Multilevel Logistic Regression; Average marginal effects; * p<0.05, ** p<0.01, *** p<0.001; HH = household