

“What good are they going to do with our information?” - UK Citizens’ Perceptions of the 2011 Census

Miguel Malheiros, Charlene Jennett, Sacha Brostoff, M. Angela Sasse
Department of Computer Science
UCL

The census aims to collect a snapshot of the UK population to better plan public services. It asks for a wide range of questions which citizens may perceive as invading their privacy. Privacy concerns may lead to census undercount and unfair allocation of resources. We describe two studies on UK citizens’ perceptions of the 2011 UK census. In Study 1 we interviewed 11 participants to identify several privacy concerns. In Study 2 we surveyed 174 respondents to delve deeper into the perceived sensitivity of census items and the prevalence of privacy protection behaviours – such as falsifying or omitting data - in census respondents. We found that: (1) individuals who submit their census later; (2) individuals who are less comfortable with disclosing census data items; and (3) ethnic minorities are significantly more likely to engage in privacy protection behaviours. We recommend that census authorities should focus on addressing citizens’ privacy concerns to maximise data quality. They should aim to collect data that has a demonstrable benefit for service planning and low privacy cost for respondents.

UK Census, privacy, personal data, privacy protection behaviours.

INTRODUCTION

The 2011 UK census was sent to every household in the country and asked for details on age, job, education, ethnicity, and religion of its members among other questions. In the UK, a census of the population has been taken every ten years since 1801, with the 2011 one being the 21st. While originally only the number of births, deaths, and marriages were registered, the census evolved to collect a much larger range and quantity of data.

The main goal of the census is to obtain a snapshot of the UK population on a given day to provide a basis for planning public services such as health, education, or transport (ONS, 2013). A census undercount (missing significant numbers of responses) can lead to local populations being underestimated, which means they receive less money from central government. For instance, in the 2001 census, Westminster's population was estimated to be 181,000 - 65,000 less than the 2000 estimate - causing the council to lose 6 million pounds in annual grants from the government (Boyle & Dorling, 2004).

In 2011, census staff tried to combat non-response by identifying people who had not completed the forms. If they persisted in their refusal they could be fined up to £1000 or face criminal charges

(Ross, 2011). It was also the first year that it was possible for UK citizens to submit their census online. However, neither of these measures tackles what we believe to be the main cause of non-response: not all citizens are comfortable disclosing the personal data asked in the census.

Privacy Concerns

The census data collection has been criticised as an invasion of privacy. Even before the first UK census, one member of the House of Commons, in 1753, called the idea of taking an annual account of the population "*an interference into domestic concerns*" (The Sidney Morning Herald, 1851). More recently, privacy advocates have raised concerns that census data can be used for purposes other than the ones stated in the forms (Boyle & Dorling, 2004). The fact it contains rich details about a whole country, including small communities, makes it a valuable resource for scientists: it allows them to understand migration flows, changes in circumstances of specific populations, or estimate birth, death, and disease rates. But census data is also employed by marketing companies to profile individuals according to their geo-demographic details (e.g. Mosaic UK; Experian, 2013). This can result in social sorting, where individuals are offered

services under harsher conditions (e.g. high insurance premiums or interest loans) or simply denied services because of the way they have been profiled (Lyon, 2003).

It has been argued that these secondary uses of census data are not part of the social contract between the state and the citizen completing the form (Heeney, 2012). Individuals do not expect their details to be used in such ways and thus, the contextual integrity of the data is violated (Nissenbaum, 2004). The ONS assures individuals of the confidentiality of census data, stating that their *"personal census information is not shared with any other government department, local councils or marketing companies"* and that the data will only be used to *"produce statistics and for statistical research"* which *"will not reveal any personal information."* This assurance ignores that with modern machine learning and data mining techniques statistical data can be de-anonymised or used for profiling purposes that have real consequences for individuals (Heeney, 2012). One does not have to go further than the stated goal of the census, making public funding decisions, to see this. While based on statistical data, decisions on how much money to grant each council will undoubtedly affect the citizens that live there.

Previous Research

Research into the factors that influence census response rates has mainly been conducted in the US. A low-level of response to the 1990 US census led the Bureau of the Census to commission a study into the causes of non-response (Singer et al., 1993; Couper et al., 1998), which revealed that privacy and confidentiality concerns significantly affected the likelihood of submitting the census forms. The authors defined confidentiality as keeping data given to one receiver inaccessible to other receivers and privacy as keeping personal data inaccessible to others in general. They also found race to be a significant determinant for response behaviour: black non-Hispanic individuals were significantly less likely to submit their census than white non-Hispanic or Hispanic individuals. The study was repeated for the 2000 US census with similar results (Singer et al., 2003): the concern that census data could be misused was again a predictor of response behaviour. Furthermore, the length of the form participants received at home (in the US there is a long and a short version of the census) had a significant effect on response: individuals who received the short form were more likely to submit it compared to the ones who received the long version.

Privacy concerns were also behind the census boycotts in Germany in the 1980s (Efferink, 2012). Vague statements about sharing data between different government organisations in the census

guidelines led to activists protesting against the 1983 census and calling a boycott. Many young citizens had suffered harassment from security forces and were deeply suspicious of the government data collection efforts. Technological advances had also made it easier to share data between institutions. The planned boycott started to get support from other citizens. The census was eventually deemed unconstitutional by the constitutional court, which ordered the government to redesign it. It was then re-launched in 1987, only to be boycotted again.

In the UK, Simpson (2003) provides some information regarding the 1991 and 2001 censuses non-response rates. Non-response was higher among young adults or socially excluded individuals, such as people who had recently migrated, were living by themselves, or were unemployed. In terms of item non-response, the 2001 census was worse than the 1991 one, with items like employment status, qualification, and workplace address having a non-response rate between 5% and 10% across the whole country. The items with the lowest non-response rates were age, sex, and marital status (less than 1%). The different levels of non-response for the census items support current privacy theory, which states that personal data items have different levels of sensitivity which directly affect disclosure rates (see, for example: Metzger, 2007). The ONS uses several techniques to compensate for missing data, such as filling empty fields with estimates, but the quality and value of the data are still undermined.

Our Research

The consequences of both census and item non-response underline the importance of minimising negative reactions to the census. In this paper, we describe two studies that investigated people's perceptions of the 2011 UK census questionnaire – in particular, their privacy concerns. We wanted to capture citizens' opinions while the experience of filling in the census was still recent; and conducted both studies in early April - a week after the census day, 27 March, had passed. In the first study, we asked 11 participants from an opportunity sample to fill in their census forms in the presence of an experimenter while they thought aloud about their perceptions of the census questions. The interviewees were probed about their attitudes towards the different questions and the reasons for their perceptions. Interview findings informed the design of a second study, an online questionnaire inquired a national representative sample of 174 participants about their comfort disclosing each of the census items and whether they had chosen to engage in privacy protection behaviours, e.g. non-response when filling the 2011 census.

STUDY 1: INTERVIEW STUDY

Aims

We conducted a series of interviews aimed at getting an insight into how individuals perceived the UK 2011 census questions from a privacy point of view. The goal was to collect participants' overall experiences of filling in the census form (e.g.: when they submitted the form, whether they completed it online or on paper) as well as their perceptions of the value proposition of answering and submitting the census, i.e.: whether participants considered it worthwhile to complete the census given any potential privacy concerns regarding the data they were asked to disclose.

Method

11 participants were recruited via an opportunity sample (6 female, 5 male). Their ages ranged from 19 to 56 years (mean age=28 years, SD=10.38). 4 were full time students, 4 were unemployed, 2 were full-time employed, and 1 was part-time employed/student. All participants were eligible to complete the 2011 UK census. For the majority (8 participants), it was the first census they had been required to complete. One participant had also completed the 2001 census, and another could not remember whether or not she had completed a previous census.

We provided participants with a document containing the text describing the purpose of the 2011 census¹: *"A message to everyone - act now. Everyone should be included in the census - all people, households and overnight visitors. It is used to help plan and fund services for your community - services like transport, education and health. Taking part in the census is very important and it's also compulsory. You could face a fine if you don't participate or if you supply false information. Your personal information is protected by law and will be kept confidential for at least 100 years. So help tomorrow take shape and be part of the 2011 Census."*

We also provided printed copies of the 2011 census², which consists of three sections:

- (i) **Household questions:** 14 items, to be completed on behalf of all household members;

- (ii) **Individual questions:** 43 items, to be completed by each member of the household; and
- (iii) **Visitor questions:** 4 questions, to be completed on behalf of anyone visiting on the census day, Sunday 27 March 2011.

One-on-one semi-structured interviews were conducted in a lab setting. First, participants were asked if and when they had submitted the form and what their general knowledge about the census was before they filled it in. They were then asked how they felt about the census being compulsory and how important they thought it was to complete it. After, they were asked to fill in a copy of the census while *"thinking aloud"*, i.e.: voicing their perceptions of each of the census' questions. At the end of the interview participants were again asked about their general impressions of the census form and its questions, potential privacy issues, and the benefits of submitting it.

Interviews took between 30 and 45 minutes and were audio-recorded. At the end of the interview, all participants were fully debriefed and received £5 for taking part. Filled-in census copies were either taken home by the participant or destroyed.

Results

Interviews were transcribed and analysed using the thematic analysis method (Braun & Clarke, 2006) to identify passages of text which are representative of some interesting pattern, coding them in consistent fashion, and then grouping those codes in themes that help make sense of the data and answer the research questions. In our particular case, we looked for quotes that revealed the factors that influence participants' perceptions of the census in general, and of particular census questions, and we tried to understand how those factors relate to each other. We now describe six themes we discovered in the interview data.

Perceived Relevance

The most commonly expressed theme (10 participants) in the interviews was that of *perceived relevance* of a question. Participants perceived a question as relevant if they understood why it was being asked in the context of the census, and how it related to the stated aims of the census: planning and improving local community services. When participants understood the purpose of a question, they had a more positive perception of that question. For example, Participant 3 (P3), when discussing census question: "How do you usually travel to work?" said:

"I think that would be quite important. They need to know those things, for transport and that, I think that's really important." – P3

¹ This text was the same introductory text available on the 2011 Census website online submission page.

² "2011 Census questions - England." Available to download at: <http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/2011-census-questionnaire-content/index.html>

Another participant, when discussing census question: "What type of central heating does this accommodation have?" said:

"So it probably wants to get a measure of the sort of heating, and perhaps how the government can target things like loans for solar powers and things like that, trying to be more environmentally friendly. So in a way I was kind of pleased that was in there, strangely enough."
– P5

Questions not perceived as relevant were those where participant did not understand why it was being asked in this context. When this was the case, participants would question why the data receiver needed that data, and what they would do with it:

"The only problem I had was with overnight visitors. I don't know why they would count if they are just staying for a short time frame."
– P8

When they did not understand the purpose of a question, participants would sometimes advance their own interpretations for why a question was being asked:

"I don't see the relevance of this question either really. Is it to catch people out, this question's in to prove people are an illegal immigrant, I don't know."
– P5

Secondary Data Use

Six participants mentioned the data being collected would be used for purposes other than the ones stated in the census form: planning and improving local community services. Their main criticism was lack of transparency of data usage and data receiver. While four of these participants suggested ways in which census data might be abused, the two others were simply sceptical that the data would be put to good use, without pinpointing specific fears or concerns:

"What good are they going to do with our other information not related to health, education, transport? Like, what good could come out of that information really? It's only negative, if you think about it now."
– P10

Potential secondary uses of census data mentioned by participants included fighting terrorism or doing "ethnic-based" stats; checking if people were hosting lodgers and not paying tax on their earnings; or passing on health information to the NHS or health insurance companies so that they could charge more for their services. Participant statements implying data could be used for secondary purposes had a negative connotation. Participant did not mention secondary data uses they thought could be beneficial. For

example, a participant, when discussing census question 13: "How is your health in general?" said:

"Well they could use that... they could pass your information on to health insurance, and then if you want to get health insurance they might try and charge you more money" – P11

Convenience and Effort

The effort required in answering a question seemed to have an effect on how that question was perceived and how participants chose to answer, ignore, or lie. For example, questions about visitors were considered by five participants to take too much time and effort to answer because they required participants to remember if they had visitors, and to know (or find out) and fill in their details if they did. These participants admitted that they might have said they did not have any visitors even if they actually had:

"First of all, maybe I won't know all the information of theirs. And I don't think I would go the extra step of calling them and asking them for all their details. I would just leave it blank, honestly. Yeah."
– P4

Participant 5, when asked whether s/he would answer the census questions relating to visitors said:

"Probably not. Because nobody is really ever going to find out, and I don't see the point. And if you have a lot of visitors over then I really don't want to spend another half hour filling boxes [laughter]" – P7

Convenience also played a part in the format participants chose to submit their form. A majority of eight participants chose to fill in the paper version of the census. The fact that the paper census form was "right there" in front of them made it simpler for them to fill it:

"I had the paper version sent so I just filled that in, rather than get my laptop out and login... it just seemed easier to do the paper version."
– P10

"I'm quite a technical person, so you would thought that I would have gone for the computer version, but it just seemed easier. You have it [the paper form] there in front of you and you can for it at your own pace."
– P5

Sensitivity

Eight participants categorised some data items being requested as "personal" or "not personal". They were less comfortable disclosing items they categorised as personal and more comfortable disclosing non-personal items. When a data item being requested was considered too personal to disclose five participants mentioned they would equate not disclosing it or even lying:

"I think I would put no, because y'know I have had a health problem, which I think is sort of significant, but I think that is a bit too personal so I would put 'no' there." – P3

Participants seemed more likely to be comfortable with the disclosure of items not seen as personal:

"As long as nothing is personal, personal things, then I wouldn't disclose. But these things are fine." – P4

Three participants described questions as asking for "statistical" data, "demographic" data, "common" data, or "descriptive" data. All these categorisations were associated with a decreased sensitivity of the data:

"Q4 is also demographic question, so it also makes it comfortable with that." – P7

"Yeah, it's just one of those questions that's always there, in that particular order, you know 'name, gender, date of birth, address, marital status, country of birth' and so on. It's something that you just get used to filling in and you don't really think about it anymore, why it's going to be used or how it's going to be used." – P10

Contact data, on the other hand, was seen as more sensitive because it could be used to contact the respondent or his/her employer:

"Well anything on how to contact me, that I wouldn't have had appreciated." – P10

Privacy Protection Behaviours

In addition to omitting data due to the effort involved in answering, five participants also implied they might not answer or lie in some questions due to privacy concerns, such as in health or visitor related questions:

"My mum did put incorrect information on the form, because my mum thought some of the information was inappropriate. For example, how many... have any lodgers been in your house in the last two weeks, my mum felt why does she need to be telling the government this type of information? Because it's her property, she should be allowed to have there who she wants, when she wants, and not have to explain to the government why." – P11

Regarding having to provide his/her phone number participant 10 said:

"That was probably the only thing I hesitated to add. And then I just thought, should I just rip up the form and throw it away anyway?" – P10

One participant (P5) considered other respondents were likely to lie on housing and immigration-related questions for fear of the consequences. Another participant (P10) thought some people

might lie on job related questions if they were evading taxes for example.

Projected Image

The image projected by responding to a specific question in a certain way seems to impact the likelihood of the respondent actually answering. Three participants expressed that they might not have answered questions on qualifications if they thought it did not make them look good. On the other hand, they did not have any problems giving answers that portrayed them in a positive light, like the fact that they worked for a reputable company or had high qualifications:

"I think if people are not educated they wouldn't want to answer that question. I'd feel obviously comfortable to answer it, but if I wasn't educated I wouldn't want to answer it. [...] I'd feel like they might underestimate my intelligence or they might look down...." – P2

The same idea seemed to be implicit in the "nothing to hide" comments of four participants. They did nothing wrong or criminal or that makes them look bad therefore they do not mind providing the data:

"I don't think I engage in too many bad things, such criminal acts, so I don't mind disclosing all that" – P4

STUDY 2: SURVEY STUDY

Aims

In Study 1 we explored individuals' perceptions of the 2011 UK census. We identified several themes which seem to impact perceptions of the census questions and the census overall. In Study 2 we wanted to investigate further – and in a quantitative way - two of these themes. In particular, we wanted to gather *sensitivity* ratings for each of the census questions, i.e. how comfortable/uncomfortable participants felt disclosing each data item and how their perceptions could have affected their disclosure behaviour (e.g.: whether discomfort with the census could have led participants to engage in *privacy protection behaviours* or delay the return of the census form).

Based on Study 1 and past research, we generated the following hypotheses:

H1: The later participants submitted their census forms the more likely they engaged in privacy protection behaviours.

H2a: More privacy concerned individuals submitted their census later.

H2b: More privacy concerned individuals were less comfortable answering the census questions.

H2c: More privacy concerned individuals are more likely to engage in privacy protection behaviours.

H3: Individuals who are more uncomfortable disclosing census data items are more likely to engage in privacy protection behaviours.

H4: Individuals who are more uncomfortable disclosing census data items submitted the census later.

H5: Non-white individuals submitted the census later.

H6: Non-white individuals are more likely to engage in privacy protection behaviours.

Method

In April 2011, we set up our online survey with market research company e-Rewards³. Being eligible to complete the census was a pre-requisite for participation. The survey took approximately 15 minutes to fill in. Respondents were rewarded by e-Rewards for their participation. 174 UK participants (100 female, 74 male) were recruited according to a nationally representative sampling frame. Their ages ranged from 18 to 79 years (mean=46 years, $s=16.06$). In terms of ethnicity, 160 were white (92%), five were Asian (2.9%), three were Black African/Caribbean (1.7%), two were Mixed (1.1%) and four gave no answer (2.3%).

The online survey was created using the open source software Limesurvey⁴. The survey had several components. First, respondents' privacy concern was assessed using the 3-item Westin privacy segmentation index (Harris and Associates Inc. and Westin, 1998) which categorises individuals into three groups. Reliability was questionable (Cronbach's $\alpha=0.68$), which is common for the Westin index.

Second, respondents were asked whether they had completed their census form yet and, if yes, on which day did they had submitted it. Third, respondents were presented with the full list of census items (household, individual, visitors) and asked to rate how comfortable they felt disclosing each item on a 5-level scale of comfort (1=Very Uncomfortable, 5=Very Comfortable). Reliability of this scale was excellent ($\alpha=0.98$).

Finally, respondents were asked to rate to what extent they agreed or disagreed with four statements about privacy protection behaviours on a 7-level scale (1=Strongly Disagree, 7=Strongly Agree) - e.g. "To protect my privacy some questions I could have answered I did not answer at all." The four questions covered *withholding*

data, providing *incomplete data*, providing *incorrect data*, or providing both *incomplete and incorrect data*. Reliability was excellent ($\alpha = 0.90$).

Results

Three participants claimed they had not completed their census as of the date of this study. 35 participants (20.1%) answered they had completed the census on the day of the deadline: March 27, 2011. Completion dates for the census form ranged from 30 days before the deadline to 38 days after the deadline. On average, participants submitted their census seven days after the deadline.

According to the Westin Index, 40 (23%) participants were categorised as *privacy fundamentalists*, 90 (52%) as *privacy pragmatists* and 44 (25%) as *privacy unconcerned*.

Average comfort ratings with answering the census questions ranged from 4.43 to 3.54 (see Table 1), with *gender* being the item participants were most comfortable disclosing and *another address where you stay for more than 30 days a year* the least comfortable item.

When asked how comfortable they felt disclosing data about other people in their household as compared to data about themselves 63.2% of participants said they felt "*as comfortable*", while 31% answered they felt "*less comfortable*." When asked specifically how they felt disclosing data about people who had visited their household, a higher percentage of participants answered "*less comfortable*": 48.3%, while 44.3% said they would feel the same level of comfort as if disclosing data about themselves.

Regarding participants' level of agreement with whether they had engaged in privacy protection behaviours or not, 8% agreed (slightly agreed, agreed, or strongly agreed) that they had *withheld data* when answering the census. 10.3% agreed that they had provided *incomplete data*. Fewer participants, 4% agreed that they provided *incorrect data* in the census. Only 2.3% agreed they had provided both *incomplete and incorrect data*.

Privacy Protection Behaviours and Census Return Date

The *census return date* variable was measured in days away from deadline: it was positive if the participant had been late in returning the census, negative if the participant had returned the census before the deadline, and zero if the census had been returned on the day of the deadline. There was a significant and positive correlation between participants' self-reported census return date and their level of agreement on having engaged or not in each of the privacy protection behaviours:

³ subsequently re-branded as Research Now, <http://www.researchnow.com>

⁴ <http://www.limesurvey.org>

Table 1: Comfort ratings for census items

item	N	mean	s
gender	173	4.43	.801
country of birth	173	4.38	.878
language	174	4.32	.906
nr of residents	172	4.30	.912
residents	172	4.28	.927
type of accommodation	172	4.28	.847
speaks english	167	4.27	1.032
name	173	4.24	1.009
heating type	172	4.23	.888
self-contained	170	4.22	.913
nr of bedrooms	172	4.21	.906
national id	172	4.20	.979
nr of rooms	172	4.20	.911
dob	173	4.19	1.047
ethnicity	174	4.18	1.024
student	125	4.17	.957
ever worked	167	4.16	1.064
employment type	170	4.15	1.024
past address	173	4.14	1.027
marital status	173	4.12	1.063
residents' names	172	4.12	1.059
retired or student	153	4.10	1.105
own or rent	170	4.08	1.040
passports	172	4.07	1.147
qualifications	173	4.03	1.110
travel to work	161	4.03	1.069
working	169	4.02	1.091
activity business	172	4.02	1.144
job title	172	4.01	1.139
arrival in uk	97	4.01	1.066
religion	172	3.97	1.254
looking for work	151	3.96	1.064
work hours	160	3.95	1.170
supervise	165	3.94	1.188
waiting for job 2	136	3.93	1.110
waiting for job 1	142	3.93	1.115
stay uk	114	3.91	1.172
nr of cars	169	3.90	1.153
job description	173	3.85	1.267
term address	98	3.85	1.116
employer	173	3.83	1.276
support	155	3.83	1.117
residents' relationships	154	3.82	1.189
limited	167	3.81	1.264
nr of visitors	156	3.81	1.222
health	174	3.81	1.140
another address 1	141	3.81	1.201
visitors	157	3.79	1.225
employer address	164	3.70	1.331
landlord	121	3.61	1.207
another address 2	100	3.54	1.283

- withholding data ($r_s=0.145$, $p<0.05$);
- provided incomplete data ($r_s=0.128$, $p<0.05$);
- provide incorrect data ($r_s=0.150$, $p<0.05$);
- provide incomplete and incorrect data ($r_s=0.150$, $p<0.05$).

Therefore, the later participants completed their census form, the more likely they were to agree they engaged in privacy protection behaviours. **H1** was thus supported. Levels of agreement with having engaged in privacy protection behaviours were also highly significantly ($p<0.01$) and positively correlated between themselves.

Effect of Privacy Concern

There was no significant effect of Westin privacy category on census return date. **H2a** was thus not supported. There was also no significant association between Westin's privacy category and average comfort ratings. **H2b** was also not supported. The level of agreement on whether they had engaged in privacy protection behaviours such as withholding data, disclosing incorrect data, or disclosing incomplete data, was not significantly affected by Westin's privacy category. **H2c** was thus not supported.

Comfort with Disclosure and Privacy Protection Behaviours

The average comfort of participants with item disclosure was significantly and negatively correlated with their level of agreement on whether they had engaged in privacy protection behaviours:

- withholding data ($r_s=-0.371$, $p<0.01$);
- provided incomplete data ($r_s=-0.374$, $p<0.01$);
- provide incorrect data ($r_s=-0.368$, $p<0.01$);
- provide incomplete and incorrect data ($r_s=-0.363$, $p<0.01$).

Thus, participants with lower reported average comfort with disclosure of census items tended to agree more that they had engaged in privacy protection behaviours supporting **H3**.

Comfort with Disclosure and Census Return Date

The average comfort of participants with item disclosure was not significantly correlated with census return date. However, participants' census return date was significantly and negatively correlated with their level of comfort with disclosing some of the data items:

- Type of central heating ($r_s=-0.199$, $p<0.01$);
- Country of birth ($r_s=-0.128$, $p<0.05$);
- Description of national identity ($r_s=-0.129$, $p<0.05$);
- Ethnic group ($r_s=-0.176$, $p<0.05$);
- Main language ($r_s=-0.127$, $p<0.05$);

- Level of English ($r_s=-0.150$, $p<0.05$);
- Religion ($r_s=-0.140$, $p<0.05$);
- Passports held ($r_s=-0.148$, $p<0.05$);
- Qualifications ($r_s=-0.131$, $p<0.05$);
- Whether you have ever worked ($r_s=-0.131$, $p<0.05$);
- How you travel to work ($r_s=-0.146$, $p<0.05$).

H4 was thus only partially supported.

Effect of Ethnicity

When analysing the effect of participants' ethnic group on their answers non-whites were grouped together to make up for their small numbers and because we considered it relevant to investigate whether ethnic minority participants' census perceptions differed from white participants, as was observed in the US (e.g.: Singer et al., 2003).

Average census return date did not significantly differ for whites and non-whites, not supporting **H5**; however, on average, non-whites tended to agree significantly more than whites that they had engaged in privacy protection behaviours when answering the census⁵ supporting **H6**:

- withholding data ($U=284.0$, $p(\text{one-tailed})<0.01$);
- provided incomplete data ($U=400.5$, $p(\text{one-tailed})<0.01$);
- provide incorrect data ($U=348.0$, $p(\text{one-tailed})<0.01$);
- provide incomplete and incorrect data ($U=470.5.0$, $p(\text{one-tailed})<0.05$).

Moreover, for twenty-one items non-whites reported significantly lower levels of comfort with disclosure than whites:

- residents ($U=366.5$, $p<0.01$);
- nr of residents ($U=424.5$, $p<0.01$);
- residents' names ($U=396$, $p<0.01$);
- visitors ($U=377$, $p<0.05$);
- nr of visitors ($U=335.5$, $p<0.05$);
- residents' relationships ($U=295$, $p<0.01$);
- type of accommodation ($U=458.5$, $p<0.05$);
- self-contained ($U=444.5$, $p<0.05$);
- nr of rooms ($U=361.5$, $p<0.01$);
- nr of bedrooms ($U=381.5$, $p<0.01$);
- landlord ($U=256.5$, $p<0.05$);
- nr of cars ($U=364.5$, $p<0.01$);
- name ($U=517.5$, $p<0.05$);
- dob ($U=474.5$, $p<0.05$);
- another address 1 ($U=358$, $p<0.05$);
- country of birth ($U=407.5$, $p<0.01$);
- arrival in uk ($U=239$, $p<0.05$);

⁵ Tests are one-tailed since we expected non-whites to be more likely to non-respond as in the American census case.

- stay uk ($U=244$, $p<0.01$);
- national id ($U=421.5$, $p<0.05$);
- language ($U=525$, $p<0.05$);
- retired or student ($U=376$, $p<0.05$).

DISCUSSION

Due to the privacy concerns about census data raised in the first study and evidence in past research we predicted that the more people postponed the completion of the census the more likely they were to also omit or lie on their answers. This was supported by the survey study data. The later participants completed their census form the more likely they were to agree they engaged in privacy protection behaviours. This seems to indicate that more privacy conscious individuals will delay their disclosure of data given the choice. Delaying disclosure can therefore be seen as a privacy protection strategy as well. More importantly, it suggests that the later the census form is submitted the more likely it is to contain false data or omissions. To our knowledge, this is the first time this phenomenon is observed and it seems to warrant further investigation. For example, it would be interesting to determine whether data quality is high in the censuses returned before the deadline and decays steadily as more days pass. If this was confirmed to be the case then more resources could be allocated by the ONS to verify data submitted later.

We expected individuals who are more concerned about privacy, as measured by the Westin index, to submit their census later and be more likely to engage in privacy protection behaviours, but these hypotheses were not supported by the data. We also hypothesised that more privacy concerned individuals would be less comfortable disclosing individual data items, but this was also not supported. Thus, the effect of privacy concern as *measured according to the Westin index* seems improbable.

However, if we look at privacy concern as measured by the average comfort with item disclosure revealed in the survey study, then there is a significant effect on stated likelihood to engage in privacy protection behaviours. *Sensitivity* of data was also raised as an important issue by interviewees in the first study. This supports the assertion that privacy concerns can in fact negatively impact data quality and undermine the aims of the census program. In past research, sensitivity has not only been linked to privacy attitudes (Culnan, 1993; Adams & Sasse, 2011), but also to actual disclosure behaviour with more sensitive questions more likely to lead individuals to lie or omit answers (Metzger, 2007; Horne et al., 2007). Thus, addressing the privacy concerns of citizens should be a priority if the ONS wants to

maximise data quality. One possible solution for this would be to calculate the benefit obtained by each question asked in the census and compare it to the privacy cost inflicted on respondents. If an item is too sensitive and does not provide enough value, then it should be removed from the census. Another option is to make census forms shorter while complementing the data with other government sources. This is already done in countries like the Netherlands. As a side effect the census would require less effort and time from respondents.

The *effort* required to fill in the census was raised as an issue by participants in the interview study who tended to see the census as a nuisance, and not as a valuable effort that can benefit their community. Both from a privacy and usability perspective, the visitor questions in particular, seem to be seen as too invasive and requiring too much time and effort to answer. Survey findings indicate that a substantial proportion of individuals are less comfortable disclosing data about other people in their household (31%) or visitors (48%) than about themselves. Moreover, past research suggests individuals are not comfortable disclosing data about third-parties without their permission (Malheiros et al., 2012). It is unclear how beneficial, from a statistical point of view, these questions were to the ONS, so it may be advisable to remove them in future census efforts.

In the survey study, ethnicity had a clear effect on privacy concern, with non-white participants being significantly more likely to admit to privacy protection behaviours and significantly less comfortable with disclosing 21 of the census items. This supports findings of studies conducted in the US (Singer et al, 2003) where black individuals were found to be more likely to not-respond to the census. It would be important to inquire further into this issue in the UK: why do ethnic minorities in the UK feel disaffected towards census efforts and what are their reasons?

Other themes were identified in the interview study as being linked to perceptions of the census. *Perceived relevance* has been linked to privacy attitudes before (Culnan, 1993); with questions seen as less relevant in the context there they are asked being perceived more negatively. This is also the case with the *projected image* theme. As observed in other contexts, data that portrays the individuals in a bad light is usually seen as more sensitive (Jennett et al., 2012; Malheiros et al, 2012), especially when disclosed to people close to her or him (Adams & Sasse, 2001). The fact that this theme emerged in the context of the census demonstrates a potential risk for misrepresentation of respondents if they do not want the government to have a bad image of them. In fact, while census and item undercount have been looked into,

research into census misrepresentation is, to our knowledge, inexistent and this would likely be a promising avenue of research in the future. The risk of misrepresentation may be increased if they perceive that data can be transferred to organisations other than government ones, which can impact the individual in different ways. In fact, concern with *secondary data use* - i.e. data collected for the census being passed to other organisations to be processed for different purposes - has been one of the main concerns identified in citizens with regards to the census (Singer et al., 1993; Couper et al., 1998).

CONCLUSIONS

In this paper we describe two studies aimed at understanding how individuals perceived the 2011 UK census questions from a privacy perspective. Privacy concerns have been shown to affect attitudes towards the census which in turn may lead to census and item undercounts which lower the quality of the data obtained and may harm budget allocation processes. While research has been done on the determinants of census undercount in the US, little is known about this phenomenon in the UK.

Our findings suggest that (1) individuals who submit their census later; (2) individuals who are less comfortable with disclosing census data items; and (3) ethnic minorities are significantly more likely to engage in privacy protection behaviours, such as falsifying or omitting data.

While the research presented here sheds light on privacy attitudes towards the census it could benefit from a larger, and equally representative, sample. Furthermore, in addition to effort, other non-privacy issues that may influence disclosure behaviour should, ideally, have also been investigated.

The implications for practitioners are substantial. In particular, if the link between late submission of the census and decreased data quality is confirmed then any organisations that use census data must take this fact into consideration when analysing the data. The fact that some individuals - ethnic minorities in particular - are more sensitive to some of the data collection should also be addressed. Census authorities are advised to abstain from collecting data items that bring little value to the overall goals of the census and that are considered invasive and aim to collect low-sensitivity/high-benefit items instead.

Our findings also open new avenues of inquiry to other researchers. Little research has been done into the determinants of census and item undercount in the UK or misrepresentation in censuses in general. More research should be carried out to determine why ethnic minorities seem

to engage less with the census and how to tackle this issue. Also, from a usability point of view, there seems to be room for improvement. Our interviews suggest most people avoided using the online forms. Usability researchers could explore this topic further to determine the causes for this choice.

REFERENCES

- Adams, A. & Sasse, A., 2001. Privacy in Multimedia Communications: Protecting Users, Not Just Data. In A. Blandford, J. Vanderdonck, & P. Gray, eds. *People and Computers XV - Interaction Without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001*. London: Springer, pp. 49-64.
- Boyle, P. & Dorling, D., 2004. Guest editorial: the 2001 UK census: remarkable resource or bygone legacy of the "pencil and paper era"? *Area*, 36(2), pp.101–110.
- Braun, V. & Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp.77–101.
- Couper, M.P., Singer, E. & Kulka, R.A., 1998. Participation in the 1990 Decennial Census Politics, Privacy, Pressures. *American Politics Research*, 26(1), pp.59–80.
- Culnan, M.J., 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly*, 17(3), 341-363.
- Efferink, L. van, 2012. Matthew Hannah: Germany's census boycott, police tactics, oppression, biopolitics. *Exploring Geopolitics*. Available at: http://www.exploringgeopolitics.org/Interview_Hannah_Matthew_Germany_Census_Boycott_Police_Tactics_Oppression_Biopolitics_of_Populations_Nazi_Aggression_Rote_Armee_Fraktion_RAF.html [Accessed March 27, 2013].
- Experian, 2013. Mosaic UK - unique consumer classification based on in-depth demographic data. Available at: <http://www.experian.co.uk/business-strategies/mosaic-uk.html> [Accessed March 14, 2013].
- Heeney, C., 2012. Breaching the Contract? Privacy and the UK Census. *The Information Society*, 28(5), pp.316–328.
- Horne, D.R., Norberg, P.A. & Ekin, A.C., 2007. Exploring consumer lying in information-based exchanges. *Journal of Consumer Marketing*, 24(2), pp.90 – 99.
- Jennett, C., Malheiros, M., Brostoff, S. & Sasse, M. A., 2012. Privacy for loan applicants versus predictive power: Is it possible to bridge the gap? In S. Gutwirth et al. (Eds.) *European Data Protection: In Good Health?* pp. 35-52. Springer Press.
- Lyon, D., 2003. *Surveillance As Social Sorting: Privacy, Risk, and Digital Discrimination*, Psychology Press.
- Malheiros, M., Brostoff, S., Jennett, C. & Sasse, M.A., 2012. Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application. 11th Annual Workshop on the Economic of Information Security (WEIS 2012), Berlin, Germany, June 25-26, 2012
- Metzger, M.J., 2007. Communication Privacy Management in Electronic Commerce. *Journal of Computer-Mediated Communication*, 12(2), pp.335–361.
- Nissenbaum, H., 2004. Privacy as Contextual Integrity. *Washington Law Review*, 79(1). Available at: <http://papers.ssrn.com/abstract=534622> [Accessed October 3, 2012].
- ONS, 2013. Office for National Statistics. Available at: <http://www.ons.gov.uk> [Accessed March 14, 2013].
- Ross, T., 2011. Census 2011: an army of enforcers. *Telegraph.co.uk*. Available at: <http://www.telegraph.co.uk/news/uknews/8408003/Census-2011-an-army-of-enforcers.html> [Accessed March 14, 2013].
- Simpson, L., 2003. Are the census outputs fit for purpose? In *Census 2001 and Beyond* Office for National Statistics and the Royal Statistical Society Conference. pp. 11–12. Available at: http://www.ccsr.ac.uk/staff/Ludi/documents/RSSON_SNov03Simpson.pdf [Accessed March 11, 2013].
- Singer, E., Bates, N. & Van Hoewyk, J., 2011. Concerns about privacy, trust in government, and willingness to use administrative records to improve the decennial census. In *Annual Meeting of the American Association for Public Opinion Research*, Phoenix Arizona. Available at: <https://www.amstat.org/sections/srms/Proceedings/y2011/Files/400168.pdf> [Accessed March 11, 2013].
- Singer, E., Hoewyk, J.V. & Neugebauer, R.J., 2003. Attitudes and Behavior: The Impact of Privacy and Confidentiality Concerns on Participation in the 2000 Census. *Public Opinion Quarterly*, 67(3), pp.368–384.
- Singer, E., Mathiowetz, N.A. & Couper, M.P., 1993. The Impact of Privacy and Confidentiality Concerns on Survey Participation the Case of the 1990 U.S. Census. *Public Opinion Quarterly*, 57(4), pp.465–482.
- The Sydney Morning Herald, 1851. The Census of a Century Ago. *The Sydney Morning Herald*.