“You cheated on me!”
Causes and consequences of cheating in online surveys

Carol Galais (Université de Montréal) carolgalais@gmail.com
Eva Anduiza (Universitat Autònoma de Barcelona) eva.anduiza@uab.cat

Abstract

Online surveys are increasingly popular for their relatively low costs. Previous research has paid some attention to the problems of bias on samples in online surveys. However, less is known about other sources of error that may also be important in this survey methodology. Online surveys are self-administered by respondents willing to receive incentives for completing questionnaires. Thus, they may rush to finish them as soon as possible, not reading the questions carefully and/or clicking random responses. Thus, online surveys may suffer from this type of measurement error derived from cheating (answering without reading). Trap questions have become customary to identify and control for this source of error.

In this paper we analyze the causes and consequences of cheating in online surveys using an online panel survey with four waves carried out in Spain between 2011 and 2012. Our data show relatively low levels of cheating behavior, that changes along time. Age, education, time spent online, interest in politics, party closeness, thinking about the survey as a way to express opinions and conscientiousness reduce the likelihood of cheating. The consequences of cheating do not seem dramatic in our case, but vary depending on the type of question and the survey wave. Cheaters may pose a problem when focusing on factual information or building scales.

Keywords: online surveys, trap questions, reliability, survey quality

Draft version, work in progress.

Paper prepared to be presented at the Visions in Methodology workshop. Faculty of Social Science, McMaster Uiversity. Hamilton, ON, Canada. May 20-22, 2014.

1 Introduction

Online surveys are becoming increasingly popular. They have a number of advantages, notably their price when compared to surveys based on other administration modes. But also they have their own specific problems. One of these problems is that their format and administration mode makes it possible to have extreme forms of satisficing problems. Online surveys are self-administered. Respondents can rush through the questionnaire, not paying attention to what they are asked, and/or providing a random answer, failing not only to retrieve from the memory the evaluations, orientations and facts that are inquired by the researcher, but even to read the question.

Additionally, online survey respondents are typically recruited or self-selected into panels of individuals which are then sent online surveys at regular intervals and receive a small incentive for completing each questionnaire. This aggravates the problem as respondents can then turn into "professional survey takers", failing to carefully read the questions and clicking a random response, so as to quickly complete the survey and proceed to the incentive.

These disengaged respondents, with their cheating behavior pose a problem for the reliability of the variables measured in online surveys. But, why do people cheat? Do they introduce a bias? How serious is this problem? What are the consequences of cheating for survey outcomes and research based on survey data?

Trap questions are designed for the purpose of identifying cheating, so that individuals that provide random responses can be removed, weighted or taken into account in the analysis of interest. In this paper we analyze cheating in online surveys using a four-wave online panel survey carried out in Spain between 2010 and 2013. Section 2 develops the interpretation of cheating as a strong satisficing symptom. Section 3 describes our trap questions and methodology. Section 4 analyzes the individual predictors of failing to answer correctly trap questions. Section 5 assesses its consequences for subsequent analyses and the reliability of the data. Finally section 6 discusses the results and the potential strategies to deal with disengaged respondents.

2. Disengaged respondents' cheating: a "strong satisficing" problem

Online surveys present a number of specificities regarding the sources of measurement error. Some research suggests that, since there are no social cues given by interviewers, online respondents may be more likely to provide honest answers, hence reducing bias due to social desirability (Comley 2003, Duffy et al.2005). Online surveys do not have to deal with the error introduced by interviewers (Biemer and Stokes 1989, Schraepler and Wagner 2005). However, online surveys may be more vulnerable to other risks. For instance, it is very important that respondents are motivated, reassured about confidentiality, an appropriately guided through a well-designed (precise) questionnaire so that they carefully read and answer each item.

The need to solve coverage and sampling problems that affect online surveys has encouraged the use of incentives as a best practice to achieve participation in web surveys, boost response rates and decrease non-response. However, the use of these incentives may be counterproductive on respondents previously willing to take part in the survey (Zagorsky and Rothon 2008). In the words of Cobanoglu and Cobanoglu: "The incentives should also not be so valuable in price that respondents answer the survey merely to stand a chance of winning the prize. If this is the case, the results may be biased." 2003: 486).

Disengaged respondents rush to completing the survey, use less than the average number of words given to an open-ended question, falls into inconsistencies in answers to factual questions, or draws "straight-lines", this is, they give the same answer in large number of consecutive items, i.e. batteries (Herzog and Bachman 1981).

A combination of long, exhausting questionnaires, difficult questions, low levels of cognitive competence, low motivation for optimal responses and a high motivation for the completion of the survey, may derive in inaccurate responses. Seminal works on public opinion soon started to warn against non-attitudes, the existence of which they deduced from respondents' inconsistency, acquiescence or extreme volubility of opinions and judgments (Converse 1964; 1970). This is by no means a problem specific to online surveys, but with online surveys it can get extreme, as people can skip the reading and tick randomly one of the answers listed.

Satisficing theory can help us to understand this kind of problem in the context of online surveys. This theory is a contribution to the understanding of survey response problems from the perspective of bounded rationality and survey design. The term "satisficing" is a contraction of the words "satisfy" and "suffice" that denotes meeting minimum criteria for adequacy instead of optimal procedures. The term was initially used by Herbert Simon to label the mechanisms operating in decision-making process (1956). Later it was adopted to define a cognitive shortcut taken in the process of answering surveys (Krosnick 1991). Satisficing consist in giving a sufficiently good answer (that is, a verisimilar, reasonable judgment) when being asked, but skipping or disregarding some of the steps involved in the optimal answering procedure.

According to the theory of optimal answering, the process of response has four steps. In the first place, respondents are supposed to interpret the meaning on the question. Second, they should recall all relevant facts and evaluations related to the question. Third, they should integrate and summarize the information to, finally, report a summary and accurate answer (Strack, Schwarz and Wänke 1991; Tourangeau and Rasinski 1988; Tourangeau, Rips and Rasinski 2000). But inconsistencies as described by Converse, together with different biases detected and measurement errors raised serious doubts about the rigorous compliance of these four steps by respondents.

Neglecting the second and third step, or failing to follow them accurately by incurring in bias, will be a symptom of the presence of the so-called "weak satisficing" (Krosnick 1991). Weak satisficing occurs when respondents stop considering alternatives as an acceptable response has been identified, which can imply acquiescence bias, tendency to choose the first option given (primacy effects) or selecting non-opinion response options (abstention, don't know, don't answer, and so on). There is a more severe form of satisficing that happens when steps two and three are skipped altogether, this is, when the processes of retrieving information from the memory and integrating the information are avoided. This may imply endorsing the status quo instead of change options, failing to differentiate in ratings or even selecting random answers.

Online self-administered surveys provide the possibility of a yet stronger type of satisficing: respondents can skip the first step as well, fail to read and process the meaning of the question, and jump directly to a random answer. Previous research using eye traking tecniques has found that people indeed use this type of cognitive shortcuts in online surveys. Respondents pay only partial attention to the information provided in the survey (Galesic, Tourangeau, Couper & Conrad 2008). Oppenheimer et al (2009), Kapelner and Chandler (2010), and Berinsky et al (2012) find significant amounts of this kind of disengaged behavior in online surveys that, as we shall see later, can be consequential for the analysis.

The satisficing theory suggests that giving a suboptimal response to a survey question is the product of a function that considers task difficulty, ability and motivation of the respondents (Krosnick and Alwin 1987; Krosnick, Narayan and Smith 1996; Narayan and Krosnick 1996; Bishop and Smith 2001). The formula according to Krosnick (2000:7) is the following:

$$p\text{ (Satisficing)} = \frac{\alpha_1 \text{ (Task Difficulty)}}{\alpha_2 \text{ (Ability)} \times \alpha_3 \text{ (Motivation)}}$$

Thus, encouragement to think carefully about questions may increase motivation, the same way that the number of prior questions answered may cause fatigue and diminish motivation (Holbrook, Krosnick, Moore and Tourangeau 2007). Satisficing is also affected by the cognitive difficulty inherent in a question. When a question demands a hard search of memory or summarizing and ranking personal judgments is especially difficult, satisficing is more likely to occur.

On the side of the respondent, satisficing should be less likely among respondents with more cognitive skills, since they are more able to report an optimal response. Previous research in online surveys Kapelner and Chandler (2010) find that the pass rate of instrumental manipulation checks is higher among women, increases with age and education and also with some indicators of motivation (number of words of feedback).

Hence, motivated respondents are less likely to commit this form of strong satisficing. We mean by motivated those who undertake the survey with an altruistic attitude, willing to help in the research, or because they find particularly interesting the topic of the survey. Reversely, monetary-incentives oriented respondents may be more prone to rush and cheat, as the literature on incentives in online surveys suggest. Intuitively, low income may be related to material incentives to finish the questionnaire as soon as possible. Other resources, such as free time, may also be related to a higher propensity to rush through the questionnaire.

Additionally, previous research has found that some personality traits are related to higher integrity test scores (Murphy & Lee 1994) and different aspects of survey behavior (particularly non response, see Rogelberg et al 2003, Markus & Schutz 2005). For instance, acquiescence, a type of satisficing behavior, has been found to be related to agreeableness (Couch and Keniston 1960). It makes therefore sense to derive that failing to pass trap questions, as a stronger form of satisficing, can also be related to some personality dimensions. Indeed, conscientiousness has been found to be lead students to early volunteering for as research subjects (Stevens and Ash 2001, Aviv et al. 2002). This personality trait correlates intensely with another one known as “need for cognition”, held by individuals who enjoy thinking and get intrinsic rewards from mental efforts (Cacioppo and Petty 1984). Following the same train of thought, individuals scoring low in the conscientiousness dimension of personality would be more prone to disengaged behaviour in surveys. Galesic et al (2008), however, do not find any systematic differences between more and less conscientious respondents in their analysis of eye tracking. Oppenheimer et al (2008) do not find any effects of motivation (as self assessed).

In sum, answering without reading as a form of strong satisficing in online surveys will be expected to depend on the resources of the respondent (cognitive, such as formal education and online skills, or, such as free time and income), her level of motivation (interest in politics, motivations for participating in surveys) and personality traits such as conscientiousness and agreeableness.

The consequences of cheating will depend on three elements: (1) the amount of respondents that show this disengaged behavior, (2) whether or not they are different to respondents that carefully read and answer, (3) the type of analysis we want to make. If cheating respondents are not different from engaged respondents we may assume that the main consequence will be noise. This can be interpreted in different ways. Oppenheimer et al 2009 consider that the noise introduced by cheaters may affect the statistical power of the experimental design, reducing the actual number of respondents exposed to treatments. However the exclusion of cheaters may include further bias, particularly if people cheating have particular characteristics correlated with our outcomes of interest.

The type of question and analysis can also be relevant in terms of the consequences of cheating. Long batteries of behaviors used to compute indexes can provide biased results (because cheaters systematically tick in the same column). On the other hand, randomize items and reversing scales might provide a certain degree of protection against the effect of cheaters. This paper explores these possibilities carefully.

3. Data and methodology: our trap questions

Luckily, online suveys not only provide the opportunity for failing to pass trap questions, but also the possibility to detect this behavior. To identify individuals that do not read the questions of a survey as carefully as they should, researchers include trap questions. Trap questions can be considered a type of instructional manipulation checks (IMC) or screeners (Oppenheimer et al. 2009), aimed at identifying whether or not the respondents are reading carefully questions and instructions. We refer to them from now on as “trap questions” or simply “traps”.

Trap questions are not an innovation of survey research. They have been used for a long time in the arts of journalistic interviewing and oral trial proceedings, as well as in the protocols for polygraph use to detect random responses (Nye and Short 1957). They belong to the group of techniques aimed for detecting poorly engaged respondents.

Our data are taken from four waves of a five-wave panel survey[1]. In the first wave of the panel we did not include any trap question so it is not considered in the analysis. Wave 2 includes 2.226 respondents of which 836 live up to the wave 5. Previous works point to the convenience of including more than a screener in order to measure respondents' attention (Berinsky et al. 2013). The panel under study only includes a trap question per wave, but analyzing the amount of failures and hits though the panel can give a clue of the degree of change in respondents' attention along the waves.

In our case, all trap questions asked respondents, after an introductory sentence, to select a number from a scale or a particular option from a range of responses. Since this is not a real question and does not reveal any information about the respondent, an incorrect answer implies that either the respondent is clicking the wrong option on purpose, or that she did not read the question.

An incorrect deliberate answer could be interpreted as a negligent/hostile attitude towards the survey that, however, does not lead to a complete withdrawal, perhaps because of the material incentives. We consider this rather unlikely. A tick on a random answer without reading the question could be interpreted as an attempt to skip the question. Since respondents are requested to answer all questions in our survey (except for vote choice and income), ticking any answer is a way to proceed to the end of the survey quickly. Although we cannot test with our data whether a wrong answer in the trap is deliberate or just a skip, the implications are the same.

Table 1 shows the wording of our four trap questions. Note that we kept the amount of words in each trap similar and also the place of the question in the survey (they were asked as the 44th question in each survey), so fatigue cannot be a cause for differences in the number of cheaters between waves. It may, however, be related to a general fatigue with the study. Also, note that all the questions ask the respondent to select a category out of five except for the first one, where the scale has 11 values.

(Table 1 above here)

The first trap question asks respondents to select a position in the 0 to 10 scale similar to the one that is used in many previous questions of the survey. In this case 8.5 of the sample failed to click the answer requested. In the following waves, the questions require to select one of five options, ordered in a way similar to other questions. In these cases the percentages of trapped respondents are marginally smaller (6.8, 4.6 and 6.4% by wave, respectively). Hence, fatigue due to a panel effect does not seem to play a relevant role in as the three last waves have very similar levels of trapped individuals. Also, the amount of options seems to be related to a higher proportion of cheaters, as it increases difficulty, or else, it reduces the chance to hit the right option by chance. We must bear in mind than in trap questions with five response options there is a 20% chance that a cheater will tick the correct answer by chance, so cheating may be slightly underestimated here.

Table 2 explores response patterns among respondents that stay in the study for the totality of the four waves under study. This make a total of 727 individuals that were asked the four trap questions, one per wave. We see that about 87% of these resilient

[1] The first four waves of the panel have been carried out in collaboration with the CIS (studiy 2855). The fifth wave has been done autonomously by the team of the research Project "Cambio y estabilidad en las actitudes políticas" financed by the Ministerio de Innovación y Ciencia, SCO2010-18534.

respondents respond correctly to all our traps. Most of the trapped individuals are trapped only once (7.7%) or twice (3.9%). Less than 2% were trapped 3 or more times. These are small figures, if compared with other works. Berinsky et al. (2013) review the distribution of cheaters in several works and found great variability, with levels of cheating of at least 20% of the sample. Oppeheimer et al. (2009) finds failure levels of 14%. Hence, either our sample is quite diligent and conscientious, or our trap questions were quite easy to pass.

(Table 2 above here)

Table 3 depicts the different combinations of errors (X) and correct answers (V) in the trap questions. This can be helpful to see if respondents are more prone to fail in subsequent waves of the panel due to fatigue or professionalism. We cannot discard this possibility. Indeed, most cheaters were trapped only once and mostly in the first and second waves of the panel. Yet 14 people only failed in the last panel wave. Moreover, if we take into account the whole sample, we realize that failing to answer correctly our traps in former waves of the panel seems to be related with the probability of withdrawing the survey in subsequent waves.[2] In sum, this analysis of individual patterns suggest that lack of attention is not only concentrated in some individuals, but rather comes and goes along time (Berinski et al 2012).

(Table 3 above here)

4. Why do people cheat?

After previous works on satisficing revised in section 2, satisficing might be related to a lack of cognitive resources, to a scarcity of material resources –that in turn may trigger material motivations- and to low non-material resources -such as time-, as well as to a lack of interest in the topic the survey deals with (low intrinsic motivation) and to some personality trends. Hence, we handle three sets of independent variables: resources, intrinsic motivations, and personality; taking into account that among the first groups of variables, some might gauge a bent for extrinsic, material motivations to answer online surveys.

Education and time spent online will be used as indicators of cognitive resources. Formal education is an indicator of cognitive resources to process the request of information that comes with a questionnaire. Time spent online reflects online skills, which should also facilitate the process of reading and completing the questionnaire. On the other hand, it might also point to available time (when this exceeds 8 hours per day) that can be used to answer online surveys. This is especially true when it comes to time spent watching tv, which can be interpreted as an indicator of available free time (that can be spent carefully doing online surveys). Living with kids is an indicator of little time available and hence a need to rush. Additionally, low income could be interpreted as an indicator of material motivations (hence associated with more cheating). But the alternative hypothesis is also plausible, as people with lower levels of income may have more incentives to do a conscientious work and keep on being contacted for these

[2] See additional analyses on this in table A2 in the appendix.

surveys. Therefore, we expect people with high levels of all these resources (cognitive abilities, time) to be more likely to correctly answer trap questions. As for the effect of income, it can go in both ways.

We also measure motivations to answer the survey with some indicators of political involvement (not interested in politics at all 0, a little 1, quite 2, a lot 3) and party closeness (1 means close to a party, 0 not close). Moreover, we have a specific question about the reasons why respondents accepted the survey firm invitation (only asked in wave 4). The answer options range from intrinsic motivations –"I am interested in the topic"- to agreeableness -"I always answer surveys"-, including material motivations -"I did it for the points" that are later translated into prices; hence material rewards-. We will test if such declared motivations affects the impact of the aforementioned resources indicators.

Finally, we test the effect of personality on the propensity to pass trap questions, which was measured with an indicator for each of the big five personality traits (OCEAN), a battery asked in wave 3. We expect conscientiousness to be associated with lower cheating probabilities.

The hypotheses to be tested are the following:

H11. Cognitive resources (education, online skills) and free time (living without children, time spent watching tv) should increase correct answers to trap questions. In other words, those less resourceful (less educated, with less spare time) will be more prone to commit cheating.

H12. Interests in politics, closeness to parties and intrinsic (versus material) motivations for participating in the survey should increase correct answers to trap questions.

H13. Conscientiousness should increase correct answers to trap questions.

Age and sex are included as controls in the logistic regressions, where the dependent variable is a correct answer in the trap (1) versus a fail to pass the trap (0). The results are presented in table 4. We have introduced the variables in three blocks. First, socio-demographics and resources. Then, we have added motivations: interest in politics, party identification and reasons to answer surveys. The goal of the latter is detecting distaste for surveys or, contrarily, a profit-oriented profile of respondents that may rush through the questionnaire. The last group of variables gauges the Big Five personality traits. This way we seek to answer whether some individuals may be predisposed to commit this strong form of satisficing regardless their motivations and resources.

Age is positively related to giving a correct answer. Men and women, however, are equally likely to pass trap questions; which contradicts Kapelner and Chandler's findings (2010). Education is an important predictor of correct answers in all waves, though its effect seems to be reduced when we take into account motivations and personality. Income has a consistent negative effect (all waves except 4): the higher the income the less likely to pass trap questions; though its effect is only significant in wave 3. Hence it seems that low income may work as an incentive to carry out conscientious work, rather than to rush to the material compensation.

Time spent online has the expected positive effect clearly in wave 2. Then, in wave 4, its effect fades when we take into consideration personality. Spare time (as measured by time spent watching tv) only has a significant effect on the probability of giving a correct answer in waves 2 and 3, but it clearly disappears when considering motivations. Living with kids predicts has an unexpected positive effect on correct in the first wave, maybe tapping more responsible attitudes. Then its effect is never significant again.

Political orientations pointing to a higher political sophistication (and therefore to intrinsic motivations for undertaking the study) seem to consistently predict correct answers across the study: those more interested and with a party identification are more likely to give accurate answers to our trap questions. It is remarkable that the effect of interest is clearly diminished when we also consider personality. Party identification has a remarkable impact that only fades in wave 5. The battery of questions tapping the reasons to engage in this study reveals different patterns. First, opinionated people who like giving their point of view in surveys are consistently more prone to correctly answer trap questions, except in wave 5. People answering that they always engage in surveys are also more prone to accurately answer trap questions, especially in waves 2 and 4.

Finally, having interest in the topic only emerges as a significant predictor of passing the traps in the last wave, which may indicate that indeed this is a powerful reason not only for reading and answer carefully trap questions, but also to stay in a survey panel. It is noteworthy that admitting having material interests in answering the survey never reaches the level of statistical significance.

With regard personality traits, conscientiousness appears significantly and positively related to correct answers in waves 2, 3 and 4; causing a reduction in the effect of the "I always answer surveys" indicator. Agreeableness only plays a significant, positive role in the first wave of the panel, the same than neuroticism.

(Table 4 above here)

5. The consequences of cheating

The consequences of failing to ask correctly trap questions can be manifold. We analyze three possible effects. First we consider the impact of cheating on the reliability of different measures. For this purpose, we compare the cronbach's alphas and correlations of different scales, as well as test for acquiescence bias, primacy effect or straight-lining in three batteries of questions. We expect people that cheat to be less consistent, but this may depend largely on the way the questions are presented. If several items are reversed in a scale, reliability should be higher among people that pass the trap. However, in batteries of items without reverse scaling reliability could be higher among those that cheat (as they may tick systematically in the same line), showing a bias towards consistency.

Second, we test the effects of cheating on the results of a survey experiment. We expect the effect of the treatments to be larger for people that correctly pass the trap question.

Finally, we check if cheating bias the estimation of the effects of some explanatory factors. If those that pass are different than those that do not pass in terms of characteristics that are related to outcomes of interest, their inclusion/exclusion can bias the results of the analyses. We compare some explanatory multivariate analysis of turnout and participation on the whole sample without those that failed the trap question.

The hypotheses to test are the following:

H21. Cheaters are less reliable in general. They fail to report consistent factual information, and they report inconsistent attitudes.

H22. Cheating introduces noise and reduces statistical power, hence reducing the effect of treatments in experiments.

H23. When testing explanatory hypotheses, cheating biases the results towards the null hypothesis. Hence, removing or controlling for cheating should increase coefficients and reduce standard errors.

In the first place, we check the reliability of factual information facilitated by respondents. We asked the same question about the number of children they had in waves 1 and 4 of our study. Of course, some variation is reasonable. One can certainly report having more children after a year and a half of the first measure. But it is more difficult to report having less children. We have computed the number of times that a respondent reported having less children in wave 4 than in wave 1. This happened 274 times, equivalent to a 13% of the individuals that participated in both studies. Figure 1 displays the difference in the means of this variable according to their answers to our first trap question (included in wave two). Those who passed the trap question were less prone to commit inconsistencies in this factual question (10% did so), while those who failed to pass the trap question reported inconsistent answers in almost 16% of the cases, these observations being statistically significant at 95% confidence interval.

(Figure 1 about here)

With regard the reliability of attitudes, the first part of Table 5 presents the value of the cronbach's alpha for two scales measuring political participation online (6 items) and political participation offline (6 items). These two are classic batteries of questions where respondents are asked to tick whether or not they have carried out a series of participation modes. The results show clearly that trapped respondents consistently exhibit higher values in their alphas. But, instead of pointing to a higher reliability of these scales among cheaters, this seems to be reflecting that the response pattern of people that cheated is not random but systematic, following a straight line (all "yes", all "no"). Although we cannot discard that cheaters are truly more homogeneous than hiters with regards their political behavior, this will warn us against batteries of items that pave the way for straight-lining across them.[3]

[3] I refer the reader to table A1 in the appendix. Cheaters are not more politically engaged than those who pass the traps, whith the exception of online participation in waves 2 and 3. This might be pointing to their actual lower levels of political engagement, which discards any primacy or acquiescence bias among them. Also, this might be pointing to a preference for "slacktivism" than for conventional –more costly-participation. Another remarkable trend is that cheaters experience a sharp decrease in their participation rates between waves 2 and 4, maybe due to the worsening of the political and economic crisis, which could have trigger among them more fatigue and political cynism than within hiters.

A way to answer whether these higher alphas are a matter of more reliability or a symptom of unexpected consequences of cheating, is analyzing different response patterns between cheaters and honest respondents regarding reversed scales. The consistency of those that do not pass the trap should be lower for batteries of indicators in which some are reversed. The second part of table 5 shows the correlation coefficient of two pairs of indicators. Two items tap attitudes towards migrants. In one of the items the highest value in the scale 0-10 represents a positive evaluation of migrants' effect on the economy, while in the other the highest value reflects hostility to differences in culture. Hence in a consistent pattern answers to these questions will correlate negatively. Likewise, two items tap tolerant moral values. One item is measured with a scale in which high values represent tolerance to adoption by same sex couples, while in the other high values represent low tolerance to abortion. We expect that negative correlations among items is negative, and stronger for people that pass the trap

In this case, people that have passed the trap are clearly more consistent, showing a negative and significant correlation between the paired items. Positive correlations that are found among those that fail the trap, particularly for the items that tap attitudes towards migrants, imply high levels of inconsistency. Answers on sexuality are not significantly correlated among those that did not pass the trap in waves 3 and 4; and even display the correct sign in wave 5, although the magnitude of the coefficient is much less stronger than among accurate respondents. This leads us to think that maybe some keywords such as "sex", "gay" or "abortion" cached their attention and manage to produce at least one meaningful, satisficing answer. In any case, our results point that when considering indexes built with balanced items, those that fail the trap have lower levels of reliability than those that pass; while scales made from items that go in the same direction may point to a "fake" higher consistency among people that cheat.

(Table 5 above here)

Next, we proceed to examine the consequences of cheating for the analysis of survey experiments. Wave 3 of our survey included an experiment consisted of randomly exposing respondents to some messages (allegedly issued from the Indignant movement) with relation to voting in the forthcoming general election. The aim was analyzing the impact of those messages on people's perceptions about the importance of voting for being a good citizen.

A total of 1450 participants on the survey were randomly assigned to one of three groups, two of them presenting identical vignettes but with different discourses about voting. The control group (N=472) was asked to assess to what extent voting is important to be a good citizen in a scale from 0 to 10 (our dependent variable) without being exposed to any specific stimulus. The second (N=490) and third (N=488) groups were presented the same vignette of a young man easily identifiable with the Indignant. The introductory text to the vignettes was in all cases: "As you know, some people go to the polls and others do not. When asked about what to do in the next general election 20N, this young man said the following." The text in a balloon changed for each one of the two treatments. The first stimulus uses one of the most popular slogans of the 15M (they don't represent us) as a justification of abstention. The young man says: "No. I won't vote. Not even in blank. They don't represent us". The second stimulus reflects the 15M most widespread vision of the election process as exposed above. In this case the young man says: “Yes, of course I’m going to vote, if only to cast a blank ballot.” After the treatment the same dependent variable is asked.

Table 6 presents the results of the experiment for those that pass the trap, for those that fail, and for the overall sample. The group exposed to a message encouraging abstention and backed with one of the most popular 15M slogans gave slightly less importance to the act of voting (5.4 on average), but the difference with the control group is not significant as the overlapping standard errors reveal. Contrarily, the stimulus in favor of poll attendance triggered significantly more importance to the act of voting among respondents (6.3), when compared to the control group. The difference is wide enough to suggest that 15M messages only have an effect when they are in line with the social norm that encourages voting as one of the citizens’ duties.[4] Indeed, when they are exposed to a message despising the act of voting issued from the same context and sender, their perception of the importance of voting is not significantly different from those who had not been exposed to it.

(Table 6 above here)

Respondents that fail the trap exhibit lower means for the dependent variable in all groups of the experiment. This may mean either that they are less likely to consider voting as a duty, or that they have a higher tendency to select the middle option (5) in the response scale. As expected, treatment 1 has not significant effect, but treatment 2 has. Respondents that pass the trap show a significant effect of treatment 2 that increases the dependent variable in +0,7 points. Among those that fail the trap the effect is +0.9, but it is not statistically significant because of the small N in this group. It seems that cheaters have experienced some effect of the treatments in the expected direction, but we lack the statistical power to ascertain the effects of the stimuli.

People that do not pass the trap seem to be different than people that pass the trap in terms of attitudes. In addition, they seem to be treated by the manipulation. This means that cheating is not only a problem because it can introduce noise in the analysis of treatment effects (something that is not evident in our analysis). Removing them can introduce bias as we would be removing a part of the sample with distinctive characteristics. However it does not seem that in our case the removal of people that fail the trap question will introduce a “true effect” bias (the bias produced by the high levels of attention that we get in experimental settings, Berinsky et al 2012). In our case, removing the 134 people that fail the trap actually leaves the effect of the treatment unchanged and slightly reduces statistical significance. Hence, our 2.2 hypotheis is disconfirmed.

Last we proceed to estimate the effects of failing to pass trap questions with a series of predictive models of turnout. These models subsequently consider the whole sample, a sample “clean” of disengaged respondents and again the whole sample including the fact of passing/failing to answer the trap as a control. Table 7 shows the logit estimations of turnout in 2008 elections in waves 2, 3 and 4. Wave 5 was excluded because the question on past voting behavior refers to 2011 elections; hence it is not directly comparable.

(Table 7 above here)

[4] A post-hoc bonferroni t.test confirmed that only those exposed to the treatment in favor of blank voting yield significantly higher means for the importance of voting when compared to the control group.

The most frequent pattern is that by removing those that fail the trap the size of some coefficients decreases, and so does statistical significance. The reduction in noise does not seem to compensate for the loss of information. If we introduce the trap as a control variable, we see that passing the trap has a positive significant effect on turnout in waves 2 and 4; and that the presence of this variable reduces the size of the effect of a number of predictors, particularly knowledge. Model fit, in the other hand, is as good as when we consider the whole sample, if not better (see wave 4)

6. Conclusions

The analysis of the trap questions included in our online panel lead us to a series of conclusions. First, answering without reading seems to be a fairly limited problem, with significant variation across time. Attention comes and goes depending on the type of trap (more/less difficult) and that interests in the topic can predict passing traps only in the fifth wave of the panel, which points out that fatigue may also be involved in the process. Note that the detected patterns in the cheating model (table 4) for waves 2-4 fade in wave 5, meaning probably that we have retained respondents with a fair amount of conscientiousness and motivation. Passing trap questions in that setting will then be a matter of a great deal of interest in the survey topic; and failing them probably a matter of some factors not tested in our models, such as boredom, pessimism or sheer fatigue.

Second, failure to pass the traps is conditioned by resources (education, free time, time online) and intrinsic motivations (interest in politics, partisanship, willingness to give one's opinion and to answer surveys in general). Low income work somewhat anti-intuitively: those with less material resources seem more careful reading and answering the traps. Actually, one of our most remarkable findings is that material motivations (conducting the survey for a reward) does not play any rol in the probability of failing to answer trap questions.

On the other hand, conscientiousness is a personality trend that consistently has a positive impact through time in the probability of passing the traps. Hence, the elder, more educated, politically engaged –at least psychologically-, more comfortable and interested by the topic at stake and more conscientious are more likely to pass trap questions. To sum up, our first group of hypotheses finds partial supporting evidence, whith intrinsic motivations and personality traits being more important than extrinsic motivations and resources.

Third, the consequences of this can be manifold and it is difficult to establish a general pattern. In our data we have seen that cheating can pose a (relatively small) problem of both noise and bias, but not always in the expected direction. Indeed, if we are interested in factual information –not usually the goal of a public opinion survey- we will find that cheaters are less reliable. In the same vein, cheaters give less reliable answers to attitude scales. Indeed, one of the most relevant findings of the paper is that the consequences can be different depending on the type of questions the survey includes. We would expect consistency to be small among people that are not reading (ie. fail the trap). However, if the questions are worded so that a "linear response" is consistent, then we cannot discard that cheaters might systematically tick a whole line of responses without reading, producing apparently consistent scales . If items are balanced with some reverse coding, then those that fail the trap will show less consistency. Researchers should take care of this and word their questions accordingly whenever this is possible.

People that did not pass the trap are not necessarily less affected by experimental treatments. Effects where similar for those that pass and those that do not pass the trap, even though, due to sample size, in the latter case the treatment effect is not statistically significant. This suggest that our experiment design may have caught the attention even of respondents prone to cheat; hence vignettes and other striking designs may be useful for this purpose. Besides, the lower baseline that disengaged respondents exhibit (and that might be inflating the effects of the experiment among them) may point to some particularities of this subgroup –namely, their lower levels of duty- that we cannot risk to lose. Excluding those that fail the trap does not provide more efficient estimates, and we have not seen the risk of "true effect" bias if we remove them from the analysis. On the contrary, treatment effects marginally decline when the trapped are removed. What seems clear is that trapped respondents have quite different values in the dependent variable and hence removing them will introduce bias in the descriptive statistics. This is especially problematic taking into account that we have an online panel surve built with quasi-opt-in respondents. Each wave is less representative of the population than the former, hence deleting cheaters –which, moreover, most of the time answer meaningfully- further aggravates the representativeness problem.

Including or excluding those that fail the trap does not seem to lead to different results in multivariate explanatory analysis. If we look at the different estimations of past voting behavior, coefficients for the "usual suspects" are largely similar between cheaters and engaged respondents. Because removing them can introduce some bias restricting the sample –even more- to a highly motivated and sophisticated subset-, perhaps a more advisable strategy is to control for them. We have shown that correctly answering the trap questions has a significant effect on electoral turnout, especially in early waves of our panel; which translates in a better model fit. We cannot discard that failing to pass trap surveys gauges some personality aspect beyond the Big Five indicators relevant for electoral behavior, such as laziness. Additionally, when this variable is included as predictor, the explanatory power of some other factors-such as political knowledge- is reduced, which means that they were gauging the level of attention and engagement of the respondents.

In general this evidence points to the need to keep those that fail the trap question into our analysis, to avoid potential bias (see appendix for a quick comparison of those that pass or fail the traps to observe the large differences particularly on political behavior), but to carefully consider in question wording the potentiality for this kind of disengaged behavior and eventually to control for its potential effect.

References

Aviv, A. L., Zelenski, J. M., Rallo, L., and Larsen, R. J. (2002). Who comes when: Personality differences in early and later participation in a university subject pool. Personality and Individual Differences 33: 487–496.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2013). Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self□Administered Surveys. *American Journal of Political Science*. Vol 0, N.0: 1-15.

Biemer, P. P., & Stokes, S. L. (1989). The optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics*, *5*(1), 23-39.

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. Public Opinion Quarterly, 65(4), 479-505.

Cacioppo, J. T. and Petty, R. E. (1984). The need for cognition: relationship to attitudinal processes. In R. McGlynn, J. Maddux, C. Stoltenberg, and J. Harvey (Eds), *Social perceptionin clinical and counseling psychology*. Lubbock, TX: Texas Tech Press.

Cobanoglu, c. and Cobanoglu, N. (2003). "The effect of incentives in web surveys: Application and ethical considerations", International Journal of Market Research, Vol. 45, nº 4, pgs. 475-488

Comley, P. (2003) Innovation in online research – who needs online panels?, MRS Research Conference Paper (Vol. 36, pp. 615-639).

Converse, Philip E. (1964) "The Nature of Belief Systems in Mass Publics". In D. E. Apter (ed.) *Ideology and Discontent*. New York: Free Press.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue.*The quantitative analysis of social problems*, *168*, 189.

Couch, A. and Keniston, K. (1960). Yeasayers and naysayers: agreeing response set as a personality variable. Journal of Abnormal and Sociasl Psychoilogy, 60, 151-174.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*,*47*(6), 615.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-Tracking Data New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. Public opinion quarterly, 72(5), 892-913.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549-559.

Holbrook, AL, Krosnick, JA, Moore, D., & Tourangeau, R. (2007). Response Order Effects In Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes.‖ Public Opinion Quarterly, 71, 325-348

Kapelner, A., & Chandler, D. (2010). Preventing Satisficing in online surveys. In*Proceedings of*.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*(3), 213-236.

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

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. New Directions for Evaluation, 1996(70), 29-44.

Marcus, B., & Schütz, A. (2005). Who Are the People Reluctant to Participate in Research? Personality Correlates of Four Different Types of Nonresponse as Inferred from Self and Observer Ratings. *Journal of Personality*, *73*(4), 959-984.

Murphy, K. R., & Lee,S. L. (1994). Personality variables related to integrity test scores: The role of conscientiousness. *Journal of Business and Psychology*,*8*(4), 413-424.

Nye, F. I., & Short, J. F. (1957). Scaling delinquent behavior. *American Sociological Review*, *22*(3), 326-331.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Schraepler, J. P., & Wagner, G. G. (2005). Characteristics and impact of faked interviews in surveys–An analysis of genuine fakes in the raw data of SOEP.*Allgemeines Statistisches Archiv*, *89*(1), 7-20.

Simon, Herbert A. (1956) "Rational choice and the structure of the environment", *Psychological Review*, vol. 63, March, compiled in, and quoted from, de, Simon (1957: 261-273).

Stevens, C. D., and Ash, R. A. (2001). The conscientiousness of students in subject pools: Implications for ‘‘laboratory’’ research. Journal of Research in Personality 35: 91–97.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, *9*(1), 111-125.
Two ways:

Zagorsky, J. L., & Rhoton, P. (2008). The effects of promised monetary incentives on attrition in a long-term panel survey. *Public opinion quarterly*,*72*(3), 502-513.

Table 1. Trap questions by wave

Wording	Responses	Wave	N failures (%)	N right (%)	Total (%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the number two on the following scale?	Scale 0-10	2. April 2011	206 (8.5%)	2,227 (91.5%)	2,433 (100%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the category bad from the list below?	• Very good • Good • Fair • Bad • Very bad	3. October 2011	134 (6.8%)	1,845 (93.2%)	1,979 (100%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the category fair from the list below?	• Very good • Good • Fair • Bad • Very bad	4. April 2012	79 (4.6%)	1,638 (95.4%)	1,717 (100%)
Could you select the category regular from the list below in order to verify that the browser works properly and we are collecting all your answers?	• Very good • Good • Fair • Bad • Very bad	5. April 2013	113 (6.4%)	1644 (93.6%)	1757 (100%)

Table 2: Number of correct responses to trap questions for the respondents that took part in all the panel waves

	Freq.	Percent
0	1	0.14
1	11	1.51
2	28	3.85
3	56	7.7
4	631	86.8
Total	727	100

Table 3: Patterns of cheating along the four waves

Pattern	N
XXXX	1
VXXX	1
XXXV	5
XVXX	3
XXVX	2
VVXX	4
VXVX	5
VXXV	2
XVVX	5
XVXV	1
XXVV	11
VVVX	14
VVXV	5
VXVV	13
XVVV	24
VVVV	631

Table 4. Logistic estimations of correct answers to trap questions.

	Wave 2			Wave 3			Wave 4			Wave 5		
	Resources	Motiv.	Personal.	Resources	Motiv.	Personal.	Resources	Motiv.	Personal.	Resources	Motiv.	Personal.
Age	.03**	.02+	.01	.07***	.08***	.07***	.04*	.06**	.05*	.05*	.06+	.07*
	(.01)	(.01)	(.02)	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
Woman	.06	.06	.03	-.29	-.17	-.20	.08	.26	.16	-.11	-.07	-.02
	(.15)	(.19)	(.20)	(.19)	(.21)	(.22)	(.24)	(.25)	(.26)	(.32)	(.37)	(.38)
Education	.15***	.12***	.10**	.23***	.19***	.16***	.20***	.15**	.13**	.24**	.14	.17+
	(.03)	(.04)	(.04)	(.04)	(.04)	(.04)	(.05)	(.05)	(.05)	(.07)	(.09)	(.09)
Income	-.07	-.03	-.03	-.18***	-.18***	-.19***	-.01	-.04	-.04	-.07	.00	-.02
	(.04)	(.05)	(.05)	(.05)	(.06)	(.06)	(.07)	(.07)	(.07)	(.08)	(.10)	(.10)
Living with kids	.33+	.59*	.60*	.33	.29	.30	.34	.13	.13	-.09	-.17	-.17
	(.20)	(.26)	(.27)	(.26)	(.30)	(.30)	(.32)	(.33)	(.34)	(.49)	(.58)	(.59)
Time Tv	.08*	.07	.08	.09+	.07	.07	-.03	-.06	-.07	.09	.09	.11
	(.04)	(.05)	(.05)	(.05)	(.05)	(.05)	(.06)	(.06)	(.06)	(.08)	(.10)	(.10)
Time Internet	.15**	.18*	.18*	.04	.03	.00	-.03	.18+	.16	.05	.09	.12
	(.06)	(.07)	(.08)	(.07)	(.08)	(.08)	(.06)	(.10)	(.10)	(.12)	(.15)	(.15)
Interest politics		.32*	.26*		.29*	.24+		.63***	.56**		.55*	.48+
		(.13)	(.13)		(.14)	(.14)		(.18)	(.18)		(.25)	(.26)
Party identification		.40*	.45*		.77***	.83***		.67**	.75**		.34	.41
		(.20)	(.20)		(.22)	(.22)		(.26)	(.26)		(.38)	(.39)
Motiv. giving my opinion		.83**	.74**		.63*	.54+		1.01**	.94**		.48	.29
		(.26)	(.27)		(.29)	(.29)		(.36)	(.36)		(.49)	(.50)
Motiv. Curiosity		-.29	-.36		-.04	-.08		.23	.21		1.06	1.06
		(.27)	(.28)		(.32)	(.33)		(.38)	(.40)		(.78)	(.79)
Motiv.interest in topic		-.04	-.06		.01	.01		.08	.04		1.21+	1.37*
		(.24)	(.25)		(.28)	(.28)		(.34)	(.34)		(.66)	(.68)
Motiv. CIS survey		.69	.59		.88	.81		.22	.22		.11	-.11
		(.61)	(.62)		(.74)	(.75)		(.75)	(.77)		(1.06)	(1.07)
Motiv. Reward		.34	.25		.10	.03		.33	.28		.70	.70
		(.21)	(.22)		(.23)	(.24)		(.28)	(.28)		(.45)	(.46)
Motiv. I always answer		.55*	.50+		.23	.15		1.17**	1.10**		.58	.57
		(.25)	(.26)		(.26)	(.27)		(.38)	(.38)		(.47)	(.50)
Openness			-.10			-.00			-.02			.21
			(.09)			(.09)			(.11)			(.16)
Conscientiousness			.32***			.29**			.28**			.06
			(.08)			(.09)			(.11)			(.18)
Extraversion			.02			-.03			.12			-.24
			(.07)			(.08)			(.10)			(.15)
Agreeableness			.21*			.12			.14			.17
			(.09)			(.10)			(.11)			(.17)
Neuroticism			.12+			.03			-.11			.19
			(.07)			(.08)			(.10)			(.14)
Constant	-.36	-1.13+	-3.21***	-1.13+	-1.62*	-3.1***	-1.62*	-1.78*	-3.01**	-1.78*	-2.16+	-4.51**
	(.48)	(.65)	(.74)	(.65)	(.73)	(.81)	(.73)	(.87)	(.96)	(.87)	(1.31)	(1.58)
Pseudo R-Squared	.038	.078	.149	.075	.105	.145	.048	.123	.168	.063	.102	.137
Obs.	2409	1705	1705	1963	1705	1705	1705	1705	1705	855	721	721

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. All independent variables measured in w2, Except for motivations (w4) and personality (w3).

Table 5. Consistency: alphas and correlations among those that failed and pass the trap

	wave 2		wave 3		wave 4		wave 5	
	Failed	Pass	Failed	Pass	Failed	Pass	Failed	Pass
Political participation online scale (6 items)**	0.83	0.67	0.86	0.68	0.77	0.67	0.78	0.71
Political participation offline scale (6 items)**	0.83	0.6	0.82	0.60	0.74	0..62	0.76	0.65
Issues migrants (2 items)	.39*	-.20*	.12*	-.21*	.23*	-.23*	..14*	.-.12*
Issues tolerance (2 items)	.14*	-12*	-.09	-.38*	.04	-.36*	-.16*	-.41*

**Alphas * Pearson correlations significant at 95% C.I.

Table 6. Average "importance of voting to be considered a good citizen" by treatment and by failing/passing the trap.

	Pass	Fail	All
Control	5.7	4.9	5.6
T 1	5.4	5.3	5.4
T 2	6.3*	5.8	6.3**
Overall F	**	-	**
N	1844	134	1979

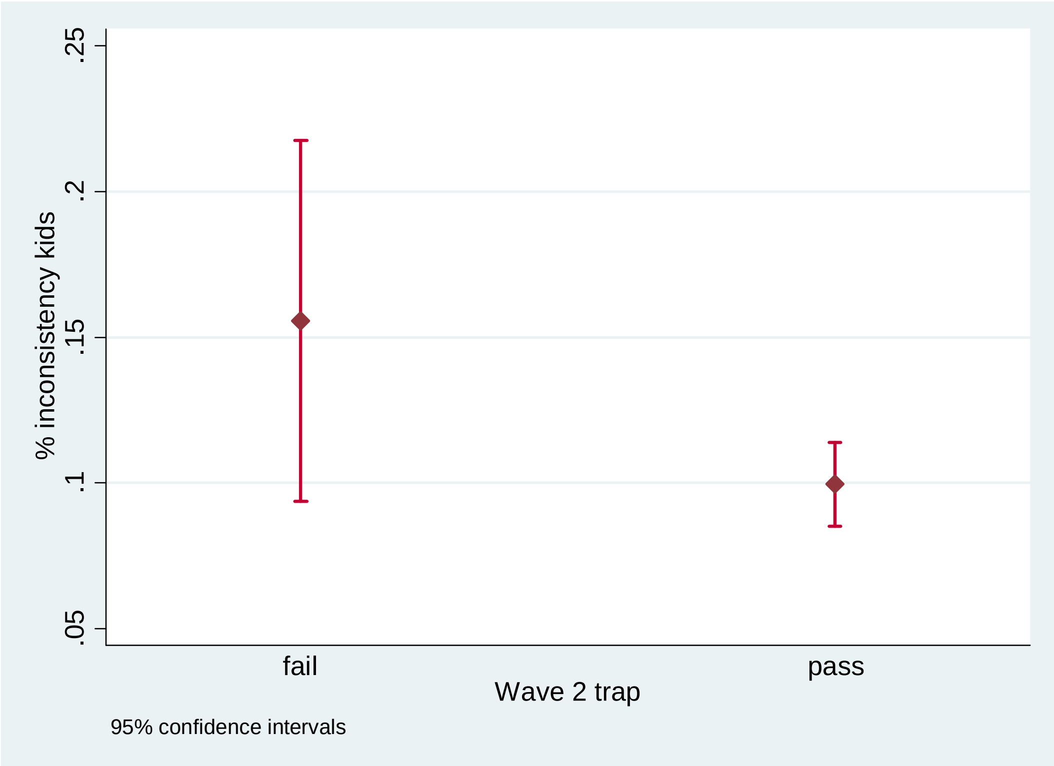
** p<0.001, * p<0.01, + p<0 Cell asterisks indicate significant differences with the control group.

Table 7. Logistic regression of electoral participation in the 2008 elections (full sample and only individuals that pass the trap question).

	Wave 2			Wave 3			Wave 4		
	(All)	(Pass)	All, trap as control	(All)	(Pass)	All, trap as control	(All)	(Pass)	All, trap as control
Age	.09**	.09*	.09*	.08**	.07**	.07**	.07**	.07**	.07**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Woman	.24*	.19+	.23*	.31*	.3*	.32*	.38*	.37*	.36*
	(.11)	(.11)	(.11)	(.12)	(.13)	(.12)	(.13)	(.14)	(.13)
Education	.11**	.13**	.11**	.11**	.11**	.1**	.08**	.09**	.08**
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)
Income	-.05+	-.05	-.05+	.01	.02	.02	-.03	-.015	-.025
	(.03)	(.03)	(.03)	(.03)	(.04)	(.03)	(.04)	(.04)	(.04)
Living with kids	.04	.11	.035	.13	.12	.11	.25	.21	.22
	(.13)	(.14)	(.13)	(.15)	(.16)	(.15)	(.17)	(.18)	(.17)
Tv time	.01	-.001	.01	.00	.01	.001	-.03	-.02	-.02
	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)
Internet time	-.01	-.04	-.015	-.02	-.04	-.022	.01	-.002	-.003
	(.04)	(.04)	(.04)	(.04)	(.05)	(.04)	(.04)	(.05)	(.05)
Interest in politics	.5**	.51**	.5**	.56**	.54**	.54**	.47**	.43**	.44**
	(.07)	(.08)	(.07)	(.08)	(.08)	(.08)	(.09)	(.09)	(.09)
Party identification	1**	1**	1**	.89**	.91**	.87**	1.03**	1.09**	1.01**
	(.11)	(.12)	(.11)	(.13)	(.13)	(.13)	(.13)	(.14)	(.13)
Knowledge 1	-.04	-.01	-.03	.07	.06	.06	.4*	.4*	.4*
	(.13)	(.13)	(.13)	(.12)	(.13)	(.12)	(.15)	(.16)	(.16)
Knowledge 2	.38**	.35*	.35*	-.02	-.01	-.02	.47**	.43**	.41**
	(.11)	(.11)	(.11)	(.18)	(.20)	(.18)	(.14)	(.15)	(.14)
Knowledge 3	.24*	.23*	.22*	.59+	.34	.53	.18	.18	.15
	(.1)	(.1)	(.1)	(.34)	(.4)	(.21)	(.13)	(.13)	(.13)
Trap (passed)			.28+			.31			.96**
			(.17)			(.21)			(.27)
Const_	-4.3**	-4**	-4**	-4.2**	-3.8**	-4.3**	-3.6**	-3.6**	-4.4**
	(.37)	(.39)	(.39)	(.5)	(.56)	(.51)	(.45)	(.47)	(.51)
R^2	.16	.16	.16	.15	.13	.15	.16	.15	.17
N	2409	2054	2409	1961	1828	1961	1707	1629	1707

Standard errors in parentheses. $^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$
Analyses are restricted to waves 2 to 4 because the question on past voting behaviour in wave 5 refers to the 2011 election.

Figure 1.. Reliability of factual information. % of respondents that reported less kids in wave 4 than in waves, by Wave 2 trap question.



Differences in means are significant at 95% level.

Appendix

A1. Mean values of political involvement by passing/failing the trap

	Passed	Failed
Voted in 2008 (wave 2)	77	57
Voted in 2008 (wave 3)	79	56
Voted in 2008 (wave 4)	79	33
Offline participation scale (wave 2)	1.7	1.7
Offline participation scale (wave 3)	1.6	1.45
Offline participation scale (wave 4)	1.8	0.9
Offline participation scale (wave 5)	2.1	1.4
Online participation scale (wave 2)	1.3	1.6
Online participation scale (wave 3)	1.36	1.39
Online participation scale (wave 4)	1.54	0.85
Online participation scale (wave 5)	1.6	1.3

A2. Relationship between passing a trap and withdrawing in the subsequent wave (pearson correlation coefficients)

	Withdrew wave 3	Withdrew wave 4	Withdrew wave 5
Passed Trap 2	-.01	.005	-.02*
Passed Trap 3	-	-.02*	-.04*
Passed Trap 4	-	-	-.05*

A3. Sample structure.

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Main sample	2100	1813	1514	1322	912
Refreshment wave, low studies	-	620	465	395	381
Recovery of participants who dropped in previous waves					464
Total N	2100	2433	1979	1717	1757
Fieldwork dates	Nov. 17 – Dec.10, 2010	May 11- May 25, 2011	Nov. 9- Nov. 18, 2011.	May 11- May 30, 2012.	May 17- June 4, 2013. - Refreshment and recovery: 16-27 October 2013.