

# Inconsistent respondents and sensitive questions

Rubén Castro<sup>1</sup>

*Abstract.* Little attention has been put on the respondents as the main cause of measurement error, regardless of the question being asked, the interviewer effect, or other factors that influence survey quality. Some individuals simply provide worse answers than others, a behavior called *satisficing*. There is much to be discovered about the consequences of satisficing on the quality of answers. An especially troublesome case is that of sensitive questions, where misreporting is quite common. In this study I conduct an empirical analysis of the effects of satisficing on answers to sensitive questions. I find a significant role for individual-level time-consistency as a predictor for answers to sensitive questions related to children, sex, condoms and HIV; Couple's reports, where couple-level fixed effects can be ruled out, confirm this result. Therefore, it is possible that satisficing has an impact on the way individuals "lie" on sensitive questions.

Keywords: Survey research, couples' inconsistencies, data quality.

## 1. Introduction

It is well known that there are several circumstances that might lead to the presence of measurement errors in survey reports (see an exposition by Biemer and Lyberg 2003), including, for example, the particular time and place where the survey is carried out, the

---

<sup>1</sup> Department of Economics, Universidad Diego Portales. I am indebted to Susan Watkins for advice, revisions and encouragement. I also thank Jerry Jacobs and Andrew Fenelon for useful advice.

ordering and placement of the questions, the characteristics of the interviewers, and the subjectivity of the questions, among many other factors (Sudman 1996, Tourangeau et al. 2000, Weinreb 2006)<sup>2</sup>.

Comparatively, little attention has been put on the respondents as the main cause of measurement error, regardless of the questions being asked, the interviewer effect, and other factors that influence survey quality. However, there is evidence that some individuals just provide better answers than others. Three examples are: Singer et al. (2000) finding that excessive efforts at recruiting respondents will bring into the sample respondents who are careless or indifferent in answering questions, ultimately damaging the quality of the information obtained, Schwarz and Clore's (1996) discussion on how optimistic answers are found among individuals receiving incentives to participate in a survey, and the Oppenheimer et al. (2009) proposal that eliminating participants who fail a quality test (around 40% of respondents on their studies) will increase the quality of the data.

### *Satisficing*

One way to look at individuals that provided answers of lower quality is to understand the process of answering as one that requires cognitive resources. The modern literature uses four stages to explore this process (Tourangeau et al. 2000)<sup>3</sup>, and all these stages require cognitive resources. Some individuals might not have, or might be less motivated to invest, those resources, something that can be called *satisficing*. Simon (1957) first introduced the

---

<sup>2</sup> Bertrand and Mullainathan (2001) argue that perhaps the most devastating problem with subjective questions is the possibility that attitudes may not “exist” in a coherent form [page 2].

<sup>3</sup> The four stages are: comprehension of the question, retrieval of relevant information, construction of judgments, and finally, selection and reporting of an answer.

term satisficing to refer to less-than-optimal decision making processes, and survey researchers applied it to answering behaviors where just a somewhat acceptable answer is provided, instead of the most thorough, true and accurate answer (Krosnick and Alwin 1987).

Satisficing could introduce random and non-random error. One satisficing-answering behavior is *rounding*, which introduces noise but not necessarily bias. Another satisficing behavior is *don't-know*, meaning the tendency to answer “don't know”, which might introduce bias due to some relevant unobserved characteristics, though Krosnick et al. (2002) find it does not. An additional satisficing behavior is *acquiescence*, meaning the tendency to answer yes; this by construction biases the answers towards yes. Another satisficing behavior is *first choice*, meaning the tendency to answer the first answer from a list of possible answers, which also introduces bias towards whatever the first option was in the list.

Satisficing is an answering behavior, and as such, is a characteristic of the individual. The amount of cognitive resources invested in answering could be associated to the circumstances of the interview (motivation factor) or can be determined by the amount of resources that the individual has (ability factor, usually measured through education), as can be seen in Krosnick (1991)'s equation for the probability of satisficing<sup>4</sup>. It could be a fairly stable characteristic. Bound and Krueger (1991) compared reported earnings with actual social security records and found that individuals who misreported their income in a given year were more likely to misreport it again in the following year. Bancej et al. (2004) found

---

<sup>4</sup> Krosnick (1991)'s equation describes the probability of satisficing in a question in particular; as such, it includes the difficulty of the question.

that inconsistent responses (on mammography history) pertain to specific socio-demographic groups, which is a common result in this literature.

### *Sensitive questions*

But in a world with more and more surveys, and individuals less and less interested in answering them (De Leeuw and De Heer 2002 document falling response rates), not much is known about the consequences of satisficing on the quality of answers. An especially troublesome case is sensitive questions, where misreporting is quite common (see a review in Tourangeau and Yan 2007). Many questions have a sensitive component. For example, sex, condom, HIV and income are sensitive topics of primary interest for research in Sub Saharan countries. Although several lines of evidence show that misreporting of sensitive questions is a deliberate rather than automatic process (Tourangeau and Yan 2007), an individual's cognitive resources invested in answering (either because of motivation or ability) could make a difference. Krosnick et al. (1996) repeatedly found that education (interpreted as a measure of cognitive resources) was positively associated with the tendency to choose *status quo* answers. Holbrook et al. (2003) found that telephone respondents were more likely to satisfice, but were also more suspicious about the interview. Thomas and Frankenberg (2002) compared reported height and weight using data from the third round of the National Health and Nutrition Examination Survey and found that men overstated their height, while women understated their weight, and also found that better-educated individuals were more likely to misreport. This article focuses on the empirical effects of satisficing in answers to sensitive questions.

The point of this article is not the truthfulness of answers. Satisficing alone might divert respondents from the truth, but not in a conscious way. In conjunction with a sensitive question, satisficing might introduce truthfulness or untruthfulness. The tendency to answer yes, for example, could be a truthful tendency if the question was “has your best friend cheated on her husband?”. Likewise, the tendency to choose the first option could introduce truthfulness. The point of this article is to figure out whether the answers to sensitive questions appear to be associated with satisficing behavior, which is relevant in understanding the misreporting of sensitive question and to emphasize the importance of individual-level causes of data error.

#### *Data and measures*

I measure individual-level time consistency in relatively simple answers and use it as an indicator of satisficing. Consistency is commonly regarded as an indicator of satisficing (see a recent example in Kaminska et al. 2010). I focus on this simple measure because it arises from an observed inconsistency in the data. I test the role of this indicator as a predictor of answers to sensitive questions related to children, sex, condoms and HIV. To better control for relevant unobserved characteristics, I focus on couple’s reports, which rules out couple-level fixed effects, and analyze wealth reports, because wealth has a somewhat less abstract meaning and it is a fairly constant variable.

The study is focused on data from the Malawi Longitudinal Study of Families and Health (MLSFH, <http://www.malawi.pop.upenn.edu>), which conducted waves of surveys in 1998, 2001, 2004, 2006, and 2008, with response rates of 78.6%, 72.1%, 67.0%, 67.9% and 67.4%,

respectively. In Section 2 I compute the consistency score. In Section 3, I explore the association of consistency with some sensitive questions. In Section 4 I explore the association of the individual's consistency score with the couples' inconsistency in reporting household wealth. The last section discusses the results.

## 2. A consistency score

I selected a subset of MLSFH survey longitudinal-questions for which answers should stay constant or vary as expected over time. The consistency score is defined as the proportion of times where the answers to these questions stay constant, or vary as expected over time. There are three steps in computing the score. First, the relatively simple questions that are asked in equivalent wording in two adjacent survey waves (1998-2001, 2001-2004, 2004-2006) are identified. Each selected question in a particular pair of adjacent waves is referred to as a “checkpoint”. Second, for each checkpoint, a dichotomous variable reflecting whether or not the respondent provided a congruent answer is computed. Third, the consistency score is simply the average of this dichotomous variable<sup>5</sup>, i.e., the consistency score is defined as the individual-level proportion of simple questions that are answered consistently across the survey's waves. This score is used to keep assumptions and complexity at a minimum. A simple score, however, is a sufficient statistic for a latent trait (Lord and Novick 1968). Leeuw and Hox (1994) used a measure called *person fit index* to

---

<sup>5</sup> Another option is to focus on all the answers to the same question over time, so that each question has the same weight in the consistency score. However, following the answers to the same question over time leaves the researcher without a clear indication of the “right” answer, when the question is answered three or more times. By focusing on checkpoints instead, the score plainly reflects the act of changing an answer from wave to wave.

explore whether consistent respondents were consistently inconsistent, and conclude that *person fit indices do not seem to be a useful tool for measuring pure respondent error*. Van den Wittenboer et al. (1990) is among the few studies that explore additional measures for the analysis of erratic response patterns.

The questions used to compute the score are: year born, ever been to school, year of first marriage, when that marriage ended, ever been outside the country, number of children, languages the respondent speaks, tribe the respondent belongs to, age of the first spouse, age of the last spouse, number of marriages, year of first marriage, year of last marriage, year of divorce/widow, main reason for divorce, ever been tested for HIV, ever receive the results of an HIV test, ever use any method of family planning, ever hear a talk at a clinic, ever hear a radio show about family planning, and number of people/relatives known to the respondent who have died from HIV.

Having selected the questions composing the score, the checkpoints are implicitly defined as described above. Each checkpoint is assigned a “1” or “0” depending on whether the answers were congruent or not<sup>6</sup>. Then this dichotomous variable is averaged across all checkpoints to produce the “consistency score”. Checkpoints are not weighted by relevance or difficulty level to keep the analysis as transparent as possible.

[Table 1 here]

---

<sup>6</sup> Not in all cases does congruent mean answering the same answer.

### *Checkpoints by wave and gender*

Table 1 shows the number of checkpoints in each adjacent wave. Women have slightly fewer checkpoints. The initial sample in 1998 consisted of approximately 1,500 ever-married women of 15 to 49 years of age, and of her husband if she was currently married, which add approximately 1,100 spouses to the sample<sup>7</sup>. Husbands in rural Malawi are typically older than their wives, and there was no age restriction on men. Indeed, men are older than women on average (37 and 31 years old respectively in 1998, 40 and 35 in 2001, and 42 and 34 in 2004). Moreover, divorce is common, and men are more likely to subsequently remarry, and remarry more rapidly, than are women (Reniers 2003). Thus, men are more likely to be older, married and to have more children, which increase the number of available checkpoints.

### *Consistency by wave*

Average consistency is higher for both genders in 1998-2001 than in subsequent waves. In 2004, the questionnaire became substantially longer, biomarkers were collected, and a new sample of adolescents was added. The longer questionnaire in 2004, as well as in 2006, increased the cognitive burden on respondents. The addition of HIV tests not only complicated fieldwork, but may have increased respondent anxiety.

---

<sup>7</sup> Subsequent waves included new spouses (2001), adolescents (2004), and the spouses of married adolescents (2006).



### *Consistency by age, education and gender*

Education appears to be associated with higher consistency, as is typical of studies in this field (Figure A). Women demonstrate higher levels of consistency than men (Figure A and B); this result is found in numerous studies: on divorce history (Mitchell 2010), income reports (Bound and Krueger 1991), drug usage (Richardson et al. 2006) and early sexual experiences (Rosenbaum 2006). Consistency appears slightly lower in older respondents (not shown); however, Reimondos et al. (2011) mention that there is mixed evidence in the literature about the effect of ageing on the quality of autobiographical memory.

[Figure A here]

[Figure B here]

### *Checkpoint bias*

Because the number and type of checkpoints is not equal across individuals, selectivity issues might introduce bias in the results. However, individuals with fewer available checkpoints, who are more likely to not be currently married, show the same average consistency (see Figure B), and the regression analyses conducted below are not affected by including the number of checkpoints as a regressor. Checkpoint bias is addressed in three ways. First, the main statistical analysis of this study is repeated, including additional dummy variables, representing the kind of checkpoints available to each individual, and the results remain the

same. Second, by aggregating checkpoints into one single score, opposing biases may cancel themselves. Third, all the regression analyses conducted in this study include controls for age, sex, education, and spouse's education, which might attenuate the influence of biases.

### **3. The association between consistency and sensitive questions on a survey in rural Malawi**

In this section I regress the answers to different sensitive questions on the consistency score. All data comes from the Malawi Longitudinal Study of Families and Health (MLSFH). Because sensitive questions typically have a strong gender component, the answers of men and women are analyzed separately. All regressions have the individual's consistency score as the key independent variable and included controls are: sex, age, education and spouse's education in years. Spouse education is an important control for women because a woman's education alone proved to be insufficient to control for unobserved socioeconomic factors. In the case of single women, this variable is replaced by the mean of spouse education and a dummy is added. If the particular question being analyzed was asked in more than one wave, wave dummies are included. Robust standard errors are used.

The sensitive questions selected cover children, condoms, HIV and sexual behaviors. In most of them, the consistency score appears to be relevant. All results are shown in Table 2.

*Children* (Regressions 1 and 2): In this case, the results show that consistent men are likely to report fewer children ever born, and fewer children ever surviving, whereas consistent

women are likely to report more children ever born and surviving. The consistency score appears very influential on these questions. *Condoms* (Regressions 3, 4, 5 and 6): Consistent men report lower condom use within marriage. Consistent women appear more likely to report “ever using condoms”, “current condom use”, and “best friend uses condoms”. *HIV* (Regressions 7, 8, 9 and 10): Consistent men and women report higher number of HIV related deaths and higher number of relatives dying because of HIV. Consistent men and women also report lower self assessed risk of being currently infected with HIV and being less worried of being infected by it in the future. *Sexual behavior* (Regressions 11, 12, 13 and 14): Consistent women and men report higher ages of first sexual encounter, higher likelihood of marrying their first partner (women only), fewer lifetime partners, and fewer best friends’ lifetime partners.

In some cases the association between the consistency score and the answer go in opposite directions among men and women. Likely, it is due to the complexities of the interplay between the social desirability implicit in sensitive questions and satisficing.

[Table 2 here]

One question that emerges is whether or not the consistency score is just capturing some relevant unobserved characteristic unrelated with the discussion at hand. In the next section I focus on questions that pertain to the context of a couple, and are answered by both spouses, whose answers should, in principle, coincide. Couple-level fix effects are included in the regressions. To minimize the distortion caused by differences in interpretation I focus on

wealth. Besides, wealth can be considered a continuous variable, which has the advantage of being less demanding in terms of statistical models and assumptions.

#### **4. The consistency score applied to couple's reports**

Couple questions are found to contain as much as 30% of inconsistencies (Miller et al. 2001). One possible explanation for those inconsistencies is varying interpretations of the question. Mother-daughter inconsistencies (Smith and Furstenberg 1994) and husband-wife inconsistencies (Ghuman, Lee, and Smith 2006) result from systematic differences in the interpretation of questions. Another explanation is social desirability. A study of Malawian couples concludes that discrepancies in reporting household wealth were systematic, with husbands likely to say “yes” and wives to say “no” (Miller et al. 2001). They also found the same patterns in the Kenya Diffusion and Ideational Change Project and the Malawi and Kenya Demographic and Health Surveys. The authors proposed that husbands over-report due to their desire to look successful while wives underreport to increase their likelihood of receiving help. On this study, I highlight an additional explanation: individuals differ on their consistency level.

Between the first and second waves, about 10% of monogamously married spouses on MLSFH gave discrepant responses to questions in which they were expected to agree (Bignami 2003). Between 20% and 30% were discrepant on responses to questions about the couples' current use of family planning and their discussions with each other about AIDS (Miller et al. 2001).

If the answer to a particular couple-related question is assumed to be a simple linear sum of the true answer plus a reporting error, then the husband's answer can be subtracted from the wife's answer, and the resulting expression will contain only reporting errors. This measure of couples' inconsistency, husband minus wife's answers, is regressed next on the husband and the wife's consistency score, plus some background variables. The relation is specified as

$$\Delta_{\text{couple}} = \beta_0 + \beta_h C_h + \beta_w C_w + \beta_x X + e \quad (1)$$

where  $\Delta_{\text{couple}}$  means husband minus wife's report, and  $C_h$ ,  $C_w$  refer to husband and wife consistency score, respectively.  $X$  refers to background variables. A negative value of  $\beta_h$  implies that, ceteris paribus, husband's consistency would be associated with smaller answers. A positive value of  $\beta_w$  implies that wife's consistency would be associated with smaller answers.

### *Estimation*

The MLSFH 2004 report has five groups of questions on household (HH) wealth. In all of them, the husband's report is higher than wife's report (not shown). Four are shown below; the fifth is excluded because responses were not significantly associated with the consistency score. I use linear regressions since the dependent variables are reasonably continuous.

Demographic variables are included in the regression. Controls include age of each spouse, age-squared of each spouse, education of each spouse (in years), and a constant. To control

for wealth issues that may also distort the coefficients, I included a proxy for HH wealth: the number of different kinds of crops the household produces. This variable, though significant in one of the models, did not change the coefficients and significance of the consistency score. Different wealth proxies produce similar results.

The dependent variables in the following regressions are the reported couple differences (husbands' answers minus wives' answer) in: HH spending on consumer goods (Model 1), number of HH assets such as beds and mattresses (Model 2), number of animals the HH owns (Model 3), and kilograms of crops produced by the HH (Model 4).

[Table 3 here]

To assess a possible bias originating because different checkpoints are available for each respondent, I repeated each regression in Table 3, but including extra dummy variables. These dummy variables capture the kind of checkpoints available for each individual. Results are largely unchanged and are available from the author upon request<sup>8</sup>.

The regression results show that couple's discrepancies are related in part to each spouse's consistency score. Even though couples fixed effects are ruled out, both husbands' and wives' estimated coefficients show that the consistency score is associated with lower reported

---

<sup>8</sup> As noted in the text, some checkpoints are available for some, but not all, individuals. Therefore, selectivity may influence the results. However, most selectivity is related to age and marital status, while results in Table 3 include married people, and age is included as a regressor. The robustness of the results in Table 3 in relation to the mentioned selectivity bias is analyzed by including some extra controls. The first control is the number of checkpoints. This number might be capturing some bias if those who are missing an important group of questions are different from those who are not missing them. Additionally, checkpoints are grouped according to survey wave and to the order they appear in the survey, and six dummies reflecting whether the individual has more than one checkpoint available in each group were created. Results in Table 3, repeated with extra controls, remain very similar, and are available from the author upon request.

wealth. Does this imply that inconsistent men and women have a tendency to over-report their economic conditions? As said earlier, not necessarily, because consistency is not necessarily associated with truthfulness. Turner et al. (1989) shows good consistency scores related to reported sexual behaviors but points out that *they do not, by themselves, tell anything directly about the systematic distortions that may occur in reporting.*

## **5. Discussion**

It is well known that there are several circumstances that might lead to the presence of nonrandom measurement error in survey reports, but most of what is known refers to the way individuals answer a specific question or a group of related questions, or to the influence of contextual factors. There is still much to be understood about quality of answers as a result of an individual-trait.

Some individuals, apparently, provide better answers than others in general (which is called *satisficing*), and this observed behavior or trait might influence the way people “lie” on sensitive questions. Satisficing alone introduces error and perhaps bias but in an unconscious way, and the interplay of satisficing and (the social desirability issue of) sensitive questions might introduce truthfulness or untruthfulness. The point in this article is that measurement error in sensitive questions, a huge issue for researchers focused on drugs, crime, sex and many other topics, might have something to do with satisficing, beyond the questions themselves.

While the time-consistency on relatively easy questions is not the only measure of satisficing, it is a fairly simple measure that could be used in other studies, and plainly reflects the observed fact of changing an answer over time in an inconsistent way. As such it offers a simple and valid indicator that could help the growth of literature around the impact of satisficing.

I used data from a longitudinal survey on rural Malawi. In line with other studies, women and educated individuals show, on average, higher time-consistency, a commonly used measure of satisficing. Controlling for different factors, I found that consistency is a statistically significant predictor of answers to sensitive questions in different areas, including sexual behavior, condoms and HIV. Also, while it is possible that consistency is just capturing some unobserved trait unrelated to the issue at hand, the analysis of the couple's answers on wealth, where couple-level fixed effects are ruled out, also shows that consistency is a statistically significant predictor of answers. Regarding the size of the effect, in couples where each spouse's score is different (half the couples), if each spouse were to have the same consistency score, the couples' gap on reporting the household wealth would be only 75% of what it was.

In some cases men's and women's consistency scores show associations of a different kind, which suggests that the interplay between satisficing and lying is not merely mechanic. Different studies found that education, regarded as the cognitive resources "owned" by the individual and as such a measure of not-satisficing, is associated with presenting oneself in *more* socially desirable ways. On the other hand, other studies (as Green et al. 2001 for



example) found that telephone respondents satisfice more and present themselves in more socially desirable ways.

Among the shortcomings of this article is its validity outside the data used for the analysis, which pertains to an undeveloped country. Also, questions are defined as easy (and therefore included in the score) arbitrarily. Moreover, a clearer conclusion would emerge if the “true” answer to a sensitive question were available. Missing data is another source of error, although the results seem stable as related to the kind of answers available for each individual. And finally, the association between consistency and a wider range of respondent traits is left for future research.

## References

1. Bancej, C. M., Maxwell, C. J., & Snider, J. (2004). Inconsistent self-reported mammography history: Findings from the National Population Health Survey longitudinal cohort. *BMC Health Services Research* 4 (1) p. 32.
2. Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications For Subjective Survey Data. *Social Science Research Network*, 91(2), 67-72.
3. Biemer, P.P. and Lyberg, L.E. (2003). Introduction to Survey Quality. New York: John Wiley.

4. Bignami-Van Assche, S. (2003). Are We Measuring What We Want to Measure? Individual consistency in survey response in rural Malawi. *Demographic Research*, S1(3), 77-108.
5. Bound, J., & Krueger, A. (1991). The Extent of Measurement Error In Longitudinal Earnings Data: Do Two Wrongs Make A Right? *Journal of Labor Economics*, 9(1), 1.
6. Ghuman, S., Lee, H. J., & Smith, H. (2006). Measurement of Women's Autonomy According to Women and Their Husbands: Results from Five Asian Countries. *Social Science Research*, 35, 1-28.
7. Green, M. C., J. A. Krosnick, and A. L. Holbrook. (2001). "The Survey Response Process in Telephone and Face-to-Face Surveys: Differences in Respondent Satisficing and Social Desirability Response Bias." Mimeo. <http://www.psy.ohio-state.edu/social/tch62a.pdf>, accessed June 01, 2011.
8. Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, 67(1), 79-125.
9. Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly*, 74(5), 956-984.

10. Krosnick, J. A., & Alwin, D. F. (1987). An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. (N. Schwarz & S. Sudman, Eds.). *Public Opinion Quarterly*, 51(2), 201-219.
11. Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213-236. Wiley Online Library.
12. Krosnick, J. A., Narayan, S., and Smith, W. (1996). Satisficing in Surveys: Initial Evidence. *New Directions for Evaluation* 70: 29-44.
13. Krosnick, J. A. E. A. (2002). The Impact of “No Opinion” Response Options on Data Quality. *Public Opinion Quarterly*, 66, 371-403.
14. Leeuw, E. D., & Heer, W. D. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (p. 41--54). John Wiley & Sons.
15. Leeuw, E.D. and Hox, J.J., (1994). “Are inconsistent respondents consistently inconsistent? A study of several nonparametric person fit indices”. In: *Measurement problems in the social sciences*, J.J. Hox & W.Jansen (Eds).
16. Lord, F. M., Novick, M. R., Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.

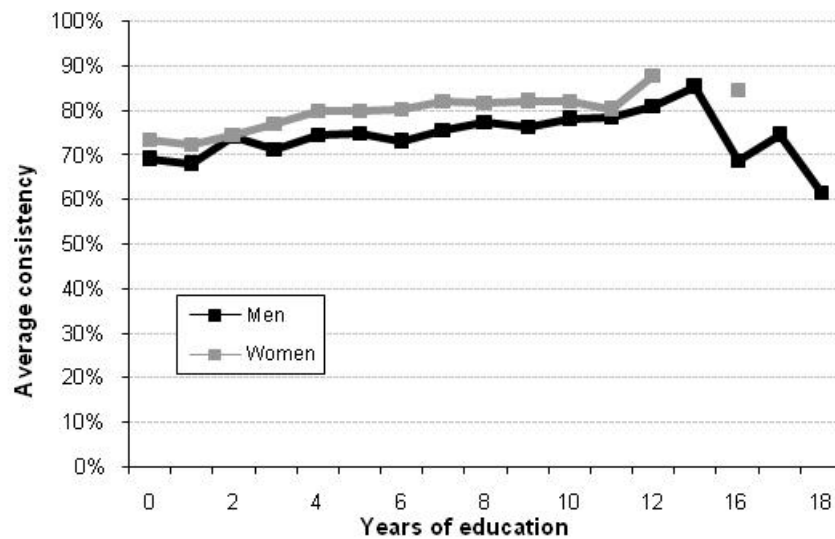
17. Mitchell, C. (2010). Are Divorce Studies Trustworthy? The Effects of Survey Non response and Response Errors. *Journal of Marriage and Family*, 72(4), 893-905.
18. Miller, K., Zulu, E. M., & Watkins, S C. (2001). Husband-wife survey responses in Malawi. *Studies in Family Planning*, 32(2), 161-174.
19. 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.
20. Reimondos, A., Evans, A., Gray, E. (2011). Reports of relationship timing: Missing data and couple agreement. *Survey Research Methods*, Vol.5, No.2, pp. 75-87.
21. Reniers, G. (2003). Divorce and Remarriage in Rural Malawi. *Demographic Research*, Special 1, 175-206.
22. Richardson, S. C., Politikou, K., Terzidou, M., Maka, Z., & Kokkevi, A. (2006). The Quality of Data Obtained from Self-Completed Questionnaires in a Survey of High School Students. *Quality Quantity*, 40(1), 121-127.
23. Rosenbaum, J. E. (2006). Reborn a Virgin: Adolescents' Retracting of Virginity Pledges and Sexual Histories. *American Journal of Public Health*, 96(6), 1098-1103.

24. Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In *Social psychology Handbook of basic principles* (Vol. 433, pp. 433-465), E. T. Higgins & E. Tory Higgins (Eds.).
25. Simon, H. A. (1957). *Models of Man*. (M. S. Archer, A. Collier, & D. V. Porpora, Eds.) *Transcendence Critical Realism and God*. John Wiley.
26. Singer, E., & Van Hoewyk, J. & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, 64(2), 171-188.
27. Smith, H. L., & Furstenberg, F. (1994). Application of a Response Model for Mother-Daughter Agreement by Race. *Social Science Research*, 23, 136-166.
28. Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass Publishers.
29. Thomas, D., & Frankenberg, E. (2002). The measurement and interpretation of health in social surveys. In *Summary Measures of Population Health*, C. J. Murray, J. A. Salomon, C. D. Mathers, & A. D. Lopez (Eds.)
30. Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response* (p. 415). Cambridge University Press.

31. Tourangeau, Roger, & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859-883.
32. Turner, C. F., Miller, H. G., & Moses, L. E. (1989). AIDS, Sexual Behavior and Intravenous Drug Use. The National Academy Press.
33. Weinreb, A. A. (2006). The Limitations of Stranger-Interviewers in Rural Kenya. *American Sociological Review*, 71(6), 1014-1039.
34. Wittenboer, G. V. D., Hox, J. J., & Leeuw, E. D. D. (1990). Chapter 14 Aberrant Response Patterns in Elderly Respondents: Latent Class Analysis of Respondent Scalability. *Social Research*, 155-162.

## TABLES AND FIGURES

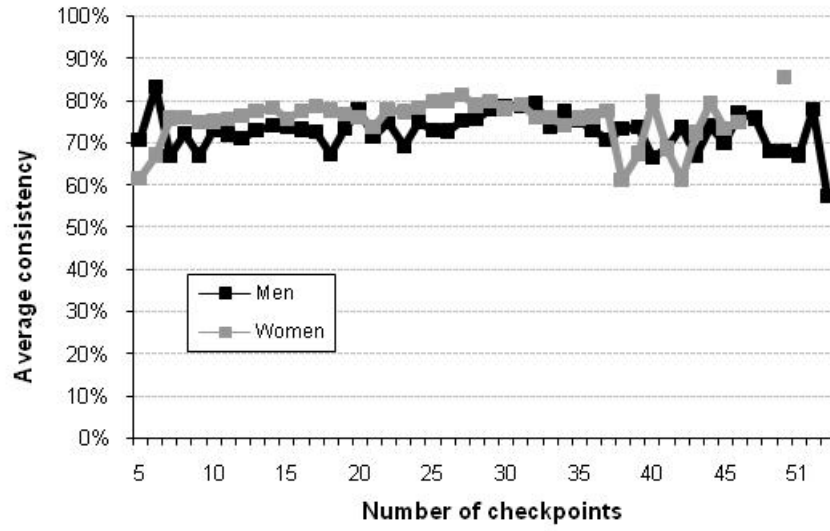
Figure A: Average consistency by education and gender  
(Education as reported in 2001. Very few individuals in the sample have more than 12 years of education.)



\* Source: Author's calculations.

\*Very few individuals in the data base have more than 10 years of education.

Figure B: Average consistency, by number of checkpoints and gender.



Source: Author's calculations.



Table 1. Consistency score and mean number of check points by wave and gender.

Waves	Mean number of checkpoints		Mean consistency score	
	Men	Women	Men	Women
1998-2001	13	11	81%	81%
2001-2004	9	6	70%	72%
2004-2006	13	12	72%	76%
TOTAL	23	20	74%	77%

Source: Author's calculations.

Table 2. Results for: Number of children ever born (1) and surviving (2), condom use in marriage (3), ever used condoms (4), current condom use (5), best friend uses condoms (6), number of people known to have died of AIDS (7), number of relatives known to have died of AIDS (8), current likelihood of being infected with HIV (9), how much they worry about being infected with HIV (10), age at first sexual encounter (11), whether the respondent married his/her first sexual partner (12), lifetime number of sexual partners (13) and whether best friend had other sexual partners (14).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Men's consistency	-1.96	-1.44	-0.96	-0.84	-1.11	-0.89	5.71	0.95	-0.86	-0.67	0.91	-0.16	-2.72	2.53
<i>P-value</i>	0.00	0.00	0.01	0.19	0.22	0.46	0.00	0.01	0.03	0.02	0.09	0.79	0.00	0.50
<i>Obs.</i>	3,360	3,253	3,344	1,078	976	352	1696	1638	2356	2356	810	876	1,656	2992
Women's consistency	0.71	1.28	-0.14	2.58	4.16	9.45	6.87	1.46	-0.38	-0.69	1.06	1.16	-2.73	-7.34
<i>P-value</i>	0.02	0.00	0.66	0.05	0.09	0.05	0.00	0.00	0.22	0.02	0.02	0.04	0.00	0.05
<i>Obs.</i>	5,180	5,032	5,047	1504	824	456	2598	2557	3510	5083	1238	1394	2,461	4566

All regressions include education, education of spouse (if married), age, age squared, single-women dummy and wave dummies. (1), (2), (7), (8), (11) and (13) are linear regressions; (3), (4), (5), (6), (12) and (14) are logistic regressions; (9) and (10) are ordered logit results; and are coded as: 0 (no likelihood) 1 (low) 2 (moderate) and 3 (high).

Table 3. OLS Regression of couple's discrepancies on husbands' and wives' consistency scores.

	Model 1: HH spending		Model 2: Number of goods		Model 3: Number of animals		Model 4: Kg of crops	
	Coef	P> t	Coef	P> t	Coef	P> t	Coef	P> t
Consistency score, husb	-1,501	0.02	-0.83	0.02	-8.67	0.03	-1175	0.01
Consistency score, wives	1,371	0.07	0.025	0.95	7.51	0.07	640	0.09
Proxy for wealth	77.00	0.23	0.04	0.14	1.03	0.01	4.13	0.88
Number of obs	651		741		355		451	

\*Model 1 included some very large, outlier numbers. In both cases, I exclude the lowest and highest 2.5% of the data. The exact questions on household spending were: Model 1: Now let's talk about household purchases made in the past 3 months. Approximately how much did your household spend in total on A) Clothes, fabric for clothes, or shoes for your children, B) School fees, school materials, or books for your children, C) Medical expenses for your children (including traditional), F) Hired labor, G) New agricultural tools/equipment and H) Large expenses related to funerals or burials for you and you family members. Model 2: Could you please tell me whether your household has any of the following? A) Bed with mattress, B) Kerosene glass lamp, C) Radio, D) Bicycle E) Pit latrine. Model 3: I'm now going to read a list of animals. Would you please tell me if your household owns any of these animals? A) Cattle, C) Pigs, D) Poultry, E) Sheep, F) Donkey, G) Rabbits. Model 4: refers to the kilograms of crops produced by the household, A) Corn, B) Soy, C) Cotton and D) Tobacco.