Experimental Tests of Survey Responses to Expenditure Questions

David Comerford
Liam Delaney
Colm Harmon

DISCUSSION PAPER SERIES

I Z A

Experimental Tests of Survey Responses to Expenditure Questions

David Comerford
*University College Dublin
and UCD Geary Institute*

Liam Delaney
*University College Dublin
and UCD Geary Institute*

Colm Harmon
*University College Dublin,
UCD Geary Institute and IZA*

ABSTRACT

Experimental Tests of Survey Responses to Expenditure Questions[*]

This paper tests for a number of survey effects in the elicitation of expenditure items. In particular we examine the extent to which individuals use features of the expenditure question to construct their answers. We test whether respondents interpret question wording as researchers intend and examine the extent to which prompts, clarifications and seemingly arbitrary features of survey design influence expenditure reports. We find that over one quarter of respondents have difficulty distinguishing between "you" and "your household" when making expenditure reports; that respondents report higher pro-rata expenditure when asked to give responses on a weekly as opposed to monthly or annual time scale; that respondents give higher estimates when using a scale with a higher mid-point; and that respondents report higher aggregated expenditure when categories are presented in a disaggregated form. In summary, expenditure reports are constructed using convenient rules of thumb and available information, which will depend on the characteristics of the respondent, the expenditure domain and features of the survey question. It is crucial to further account for these features in ongoing surveys.

JEL Classification: D03, D12, C81, C93

Keywords: expenditure surveys, survey design, data experiments

Corresponding author:

Colm Harmon
UCD Geary Institute
University College Dublin
Belfield, Dublin 4
Ireland
E-mail: colm.harmon@ucd.ie

1. Introduction

Expenditure questions are a feature of most large scale data-sets employed by economists and are intended to provide key information on the welfare of individuals and households. The data generated by these surveys form the basis of cross-sectional and longitudinal comparisons of consumption; test for the responsiveness of consumption to policy and stochastic shocks; and are used to inform theories of consumption and saving across different groups (Browning, Crossley and Weber, 2003). If measures of expenditure are biased, and more especially if bias is systematically different across groups and expenditure domains, they may lead to spurious results.

With this in mind, it is important that economists who use self-reports of expenditure develop an awareness of the potential limitations of their use. There is a well-developed literature in experimental and cognitive psychology to suggest that recall of behaviour and reporting of quantitative measures are subject to bias. Survey experiments provide a means to test for, and reveal the sources of, these biases. Ultimately the goal of this research is to develop questions that elicit expenditure as efficiently and as accurately as possible.

Experimentally testing the effects of question framing has a long history in preference and attitude elicitation (Kahneman, Ritov, Jacowitz and Grant, 1993; Diamond and Hausman, 1994; Schuman and Presser, 1996; Ariely, Loewenstein and Prelec, 2003). The results reported in these papers show that quantitative responses, which are interpreted as meaningful economic measures, are sensitive to irrelevant details of the survey process. A crucial insight of this research is that people do not have fully stable concepts of economic quantities but construct their responses when explicitly invited to do so.

This paper provides new empirical evidence on a range of potential survey effects in the context of expenditure elicitation. In particular, we address three issues of concern in survey design: 1) Question interpretation; 2) The use of features of human dialogue to address these concerns and 3) Constructed responses and response instability. The rest of this paper is structured as follows. Section 2 provides an overview of these three core concerns and the rationale for our own experiments. Section 3 outlines the experimental design and data collection methods. Section 4 provides the results of the experiments. Section 5 concludes.

2. Literature and Rationale

2.1 Question Interpretation:

A key concern when using any survey data is that the respondent and the researcher concur in their understanding of the survey question. Various mechanisms can be used to test whether this is the case. For example, Schkade and Payne (1994) used verbal protocol analysis. They asked respondents to speak aloud their thought process when responding to a willingness to pay survey. This research found that respondents paid very little attention to a crucial economic consideration, the number of birds that would be saved by their dollar contribution to an environmental project. This finding clarified the causal mechanism that lay behind previous results, which showed that responses are insensitive to the quantity of economic good being valued (Kahneman et al., 1993; Loomis, Lockwood et al., 1993).

A question that is of particular concern in expenditure surveys is the interpretation of the word "you". Previous research shows that some respondents did not recognise the distinction between "you" individually and "your household" when expressing their willingness to pay for public broadcasting (Delaney and O'Toole, 2006, Delaney and O'Toole 2008). These findings are confirmed by Lindhjem and Navrud (2008) in the context of an environmental public good. In general, the issue of how respondents interpret the word "you" in survey questions has received too little attention in the literature despite the potentially severe distortions that can result from this issue. Since this ambiguity is only likely to occur for respondents who live in households with joint finances, comparisons of single people and respondents in partnerships are likely to be biased. It may also be the case that the same respondent changes her interpretation of the word "you" depending on the domain in which it is being asked, which would complicate matters further.

2.2 Features of Human Dialogue:

With the advent of web-surveying it has become possible for surveys to monitor responses in real time and interact with respondents so as to facilitate the survey procedure. At its most basic level this makes survey response more efficient by routing respondents through items that previous responses have shown to be irrelevant. A more ambitious application is the automatic activation of a glossary of terms if there is no response within a certain time period. This strategy has been shown to increase the accuracy of response (Conrad, Schober and Coiner, 2007).

A second use of human dialogue is the "stop-and-think" prompt. Respondents who have been prompted to stop and think prior to making a judgment have been shown to attend to more, and more diverse, considerations than those who did not receive the prompt to stop and think (Zaller, 1992). The stop-and-think prompt may encourage respondents to search their memory more thoroughly for instances of expenditure on the target good than they otherwise would. Thus, the expected effect of their inclusion is to increase the amount of expenditure recalled.

2.3 Constructed Responses and Response Instability:

Many previous surveys show that preferences and willingness-to-pay can be manipulated by sometimes trivial details of survey design (Diamond and Hausman, 1994; Kahneman, Ritov & Schkade, 1999). For example, when asked the final two digits of their social security number, and subsequently asked to value a good, the money amount people give by way of a valuation anchors to the two digit social security figure (Ariely, Loewenstein and Prelec, 2003). In general, respondents do not use full and unbiased information when responding to survey questions. Instead they form their answers on the basis of information that they find most available, including features of the survey. For example, Hurd (1999) tested for anchoring and acquiescence in surveys designed to elicit the value of respondents' homes. Respondents were asked to value their home between bounds in an iterative procedure, within increasingly narrow bounds of money amounts. The experiment finds that the seemingly arbitrary choice of starting point has a significant effect on respondents' valuation of their own home.

Expenditure is potentially a more meaningful construct to respondents than their home value as they directly influence expenditure on a regular basis. Yet there is compelling reason to believe that it too will be sensitive to features of survey design. In a series of papers, Menon and co-authors have demonstrated experimentally that respondents make use of the availability heuristic to recall the frequency of behaviours (Menon, 1993; Menon, Raghubir and Schwarz., 1995; Menon and Yorkston, 2000; Raghubir and Menon, 2005). The heuristic gives fairly accurate measures when the instances of a behaviour are similar and regular (Menon, 1993). If these two conditions do not hold, however, frequency reports tend to be underestimated. Spending money is a behaviour, and so we see no reason why these insights would not apply in the context of expenditure. Indeed recall of expenditure is likely to be even more biased than recall of behavioural frequency since respondents must recall both the frequency of purchase and the amount spent. Menon's results suggest that the degree of bias

in recall will differ across domains, with infrequent purchases being understated relative to routine purchases.

There has been some previous work which supports this hypothesis. Winter (2004) randomly assigned respondents to report their expenditure on household non-durables in one of two ways: as a single aggregate figure; or as the sum of expenditure on thirty-five sub categories of household non-durables e.g. food and drink. Using the thirty-five disaggregated subcategories increased reported total expenditure and was found by cross validation with a budget survey to be more accurate. It was also found that the degree of understatement associated with the aggregated measure differs across respondent characteristics such as age. Pradham (2009, this issue) also finds that higher levels of aggregation in question elicitation yields lower aggregate reported consumption.

In another paper, Winter (2002) asked respondents to report how much they spent in total in the past month using a range card with bracketed categories. There were three conditions: one offered expenditure categories that were clustered at the lower end of the distribution so as the expenditure of the median respondent would appear relatively high. A medium treatment distributed the categories around the expected median. The high treatment offered categories that were high relative to the median of the population. Winter finds that this presentation has a significant effect on responses, with the effect most marked in the low condition.

Anchoring is not the only reason why the presentation of category brackets might impact on people's responses. People tend to avoid rating themselves at an extreme point on a distribution. Oswald (2008) demonstrates that the distribution of height across a population exhibits greater kurtosis when measured on a subjective scale than when objective metrics are used. On the non-objective scale respondents in the tails of the distribution report themselves as closer to the average than they actually are. Mid-point bias has been noted in a number of other papers (Dawes, 2000; Garland, 1991).

One reason why the mid-point bias is likely to occur in an expenditure context is that respondents infer population averages, or possibly even behavioural norms, from the presentation of the categories. For example, Haisley, Mostafa and Loewenstein (2008) demonstrate that manipulating income brackets so that the median income looks higher than it actually is increases the probability that respondents will purchase a lottery ticket. Their results indicate that the presentation of the brackets provides respondents with subjective information as to their place in the income distribution. This effect is large enough to change

behaviour, in this case to alter the respondents' choice between receiving cash and receiving lottery tickets. Specifically, more respondents choose lottery tickets when the categories are presented in such a way as to make their income appear relatively lower.

In this paeper, we further examine the extent to which the time-unit used influences the answers given. If respondents are recalling and reporting average expenditure accurately then there is no reason for the time-scale used to influence their reports. However, there is strong reason to suspect that respondents may report larger pro rata expenditure when asked to report on small time-scales. Respondents who employ the availability heuristic will find it less difficult to recall individual items over a short period, and so they will report more of them (Menon and Yorkston, 2000).

Moreover, Prelec and Loewenstein (1991) write of the "peanuts effect" whereby small money amounts are dismissed as trivial. In the context of gambling, risk aversion is lower for small amounts because small losses are predicted to make less of an affective impact (Weber and Chapman, 2005). The cumulative impact of a series of very small monetary amounts is less than the impact of an equivalent single amount (Morewedge, et al., 2007). Due to the fact that intense affective experiences are privileged in memory (Wirtz, Kruger, Scollon and Diener., 2003), small expenditures are likely to be forgotten though cumulatively they may be considerable. Also, it may be psychologically more aversive to report high absolute expenditure figures, particularly for indulgences such as alcohol.

3. Method and Participants (INSERT TABLE 1 ANYWHERE)

Participants were recruited at a bus station; a train station; on the university campus and on a commuter train travelling between Dublin city centre and various suburbs. They were asked to complete a paper survey. No monetary incentive was offered and participants were assured that the survey would take no more than five minutes. Paper surveys were randomised prior to going into the field so as to ensure that participants were randomly assigned to one of forty-eight survey conditions as follows:

Stop-think		*Item list then aggregate*		*High-scaled Brackets*		*You*		*Weekly*
	X		X		X		X	*Monthly*
No prompt		*Aggregate only*		*Low-scaled brackets*		*your household*		*Yearly*

Fig 1: A (2 x 2 x 2 x 2 x 3) survey randomisation gives forty-eight variants of the survey

To ensure the integrity of the experiments, respondents were instructed not to consult with each other or look at the questionnaires of other respondents when answering the questions. 444 respondents were recruited at bus stations, 443 at Train Stations, 170 on commuter rail lines, and 172 on the college campus.[1] 1,218 surveys were distributed over five days.

Survey experiments have the benefit that the hypothesised causal stimulus can be randomly assigned. In theory, random assignment means that respondent characteristics, both unobserved and observable, are orthogonal to the causal mechanism of interest. In practice, samples are seldom large enough to guard against coincidences. To validate the randomisation procedure, we report the results of probit estimation of respondent characteristics on survey assignment in Table 1. As can be seen, there are observable differences in the samples for both the Stop-and-Think and the Recency Christmas tests. For all other survey conditions it suffices to control for the survey condition only.

4. Experiments (INSERT TABLE 2 and 3 ANYWHERE)

In this section we report the results of the survey experiments we performed. The experiments are grouped according to the issues that they test for: Question interpretation; the use of human dialogue; constructed response. Unless otherwise stated, the regressions that follow control only for survey condition. These results are displayed in Table 2. The open-ended expenditure questions were transformed by a Box-Cox procedure so as to correct for skewness in the raw data. Because several hypotheses are being tested using the same sample, we also examined the extent to which the results are robust to the use of a standard test for multiple comparison effects, the Holm-Bonferroni adjustment (Holm 1979).

4.1 Question Interpretation & Human Dialogue – "You" and "Your Household"

Our survey uses a pen-and-paper self-completion format. Even with such rudimentary technology, however, we believe that there is scope for applying human dialogue cues to improve the accuracy of survey response. In a bid to clarify how the respondent interprets the questions eliciting expenditure on alcohol, food and drink, we ask a follow-up question. Respondents were asked whether their responses referred to their individual expenditure; their household expenditure; or a combination of the two.

---

[1] All of the results reported in Table 2 are robust to including a dummy for sampling location.

Hypothesis: Survey responses are not sensitive to the distinction between individual expenditure and collective household expenditure

Procedure: Respondents were randomly assigned to report either how much "you" spent or how much "your household" spent on motoring expenses, food, alcohol and in total on all things. Having made their report of expenditure, respondents were then asked a clarification question. The clarification question asked whether the report is the total amount spent by the individual respondent alone; the amount spent by the individual respondent and other members of their household; or, the total amount spent by the household.

Results: The results in table 3 clearly illustrate that respondents struggle to differentiate between their individual expenditure and that of their household as a whole. The results refer only to respondents who are living with at least one other person. Approximately 20 per cent of respondents interpret "you" as referring to their household when estimating their expenditure on alcohol and their total expenditure. The ambiguity is most marked when reporting expenditure on food to be consumed in the home. One third of the sample reports their household expenditure when asked to report "your" expenditure on food. Responses are just as ambiguous when the survey asks respondents to report their expenditure at the level of their household despite the fact that this formulation is less ambiguous than being ask simply "your" expenditure. Almost two thirds of respondents who are living with a partner or relatives reported their household expenditure on food when asked to do so.

Conclusion: The interpretation of the words "you" and "your household" differs across respondents. Moreover, an individual respondent will interpret "you" and "your household" differently in different domains.

4.2 Human Dialogue - Stop-and-think

Hypothesis: Respondents primed with a stop-and-think prompt will report a higher total expenditure than others because they access a wider range of considerations (Zaller and Feldman, 1992).

Procedure: Prior to answering questions on expenditure, a random subsample received the advice: "Please think in detail before answering the questions which follow as many people forget what they have actually spent". This advice is expected

to cause people to give more consideration to the question and to retrieve information that they otherwise would not. If this is so, people primed to stop-and-think will report higher expenditure than those who are not so primed.

Results: The stop-and-think prompt has no observable effect on reports of food, alcohol and total expenditure (table 2). However, its effect is likely to be strongest on responses to the question immediately before which it was placed. This was a question about motoring expenditure. Since only a fraction of respondents own a car we include controls in the model so as to control for any potential confounds. The coefficient on the stop-and-think prompt is insignificant also.

Conclusion: Respondents do not report higher expenditure when instructed to think in detail and reminded that they might forget some expenditures..

4.3 Constructed Responses

4.3.1 Disaggregated Prompts

Hypothesis: An itemised list of disaggregated motoring expenses will help respondents recall motoring expenditure that would otherwise be forgotten.

Procedure: A random sample of respondents received an itemised list of motoring expenses to aid in recall of total expenditure. We predict that respondents who receive the list will report having spent more in total than respondents who are simply asked to report the total they spent on motoring. Such a list has been shown to increase reports of expenditure on household non-durables (Winter, 2004). Menon (1993) demonstrates that respondents have particular difficulty recalling infrequent and irregular behaviour compared to behaviour conducted on a routine basis. Since some motoring expenses (e.g. vehicle maintenance) are infrequent and irregular, we believe that directly reminding respondents to include these will increase total reported expenditure.

Results: Disaggregated prompts have a significant effect on reported car expenditure with respondents in the disaggregated condition reporting significantly higher levels of expenditure. Because only a subset of the sample have a car we control for observable characteristics (n = 192; t= 2.79; p = 0.006).

Conclusion: Prompting item recall increases expenditure, consistent with the evidence that respondents have difficulty remembering all aspects of expenditure.

4.3.2 Timescale effects

Hypothesis: Respondents will report lower pro-rata expenditure as the unit of time over which they are reporting increases.

Procedure: Respondents were asked to report their expenditure on food for consumption at home, expenditure on alcohol and their total expenditure all things considered. They were randomly assigned to report these on a weekly; monthly or yearly timescale. We predict that mean expenditure per year will be less than mean expenditure per month multiplied by twelve; and even less again than mean expenditure per week multiplied by fifty-two.

Results: Controlling only for survey condition, the effect of timescale is highly significant in the hypothesised direction. As can be seen in Table 2 the effect is substantial. For example, respondents reporting on the weekly scale are fifteen per cent more likely to report spending more than 2,080 euro per year on alcohol.

Conclusion: The effect of timescale on reports of expenditure is as predicted. Pro-rata expenditures decrease as the time-unit increases for all categories.

4.3.3 Anchoring to brackets

Hypothesis: Reported expenditure is sensitive to the bracketed categories chosen by the survey designer.

Procedure: Respondents were randomly assigned to report their alcohol expenditure on one of two category scales. Condition 1 has a midpoint of forty euro and five categories (€0; €1 - €20; €21 - €40; €41 - €100; €101 +). Condition 2 has a midpoint of sixty euro and six categories (€0; €1 - €40; €41 - €60; €61 - €80; €81 - €100; €101 +). If our hypothesis is correct a greater proportion of respondents will report having spent more than €40 per week on alcohol in condition 2 than in condition 1.

Results: The probability of reporting having spent over €40 or equivalent is higher if alcohol expenditure is elicited on the higher anchored scale ($z = 2.68$; $p = 0.002$). Respondents in condition 2 were seven per cent more likely to report having spent

more that forty euro per week (or equivalent if reporting on other timescales). Somewhat surprisingly, respondents in condition 2 also report higher expenditure for total and for food. However, only the alcohol result remains significant following the Holm-Bonferroni adjustment of p-values.[2]

Conclusion: The respondents use arbitrary features of the survey question to assist them in making a response. In particular, the mid-point of the scale is used by the respondent as a guide to making response.

4.3.4 Recency bias

Hypothesis: Respondents surveyed one week before Christmas will report having spent more on alcohol and food in a typical week over the past year than will respondents surveyed three weeks after Christmas. The availability bias leads respondents to refer to recent weeks when constructing a "typical" week.

Procedure: Half of the sample answered the survey one week before Christmas. The other half was recruited in mid-January.

Results: As can be seen in Table 2, there is some evidence to support the claim that respondents' reported average alcohol expenditure differs depending on the time period. The effect of answering before Christmas is contrary to that anticipated, respondents in the pre-Christmas condition reported spending less on alcohol than did respondents after Christmas.

Conclusion: We find a small though statistically significant effect of Christmas responding on estimates of average expenditure.

5. Conclusions

This paper provides novel experimental evidence on a range of potential biases inherent in eliciting expenditure. Expenditure reports are insensitive to some relevant features of the question and sensitive to some irrelevant ones. These effects relate firstly to the fact that

---

[2] Controlling for total expenditure in the food regression completely removes the effect of the scale but has little effect on the coefficient in the alcohol regression. The use of the six-point scale is equal across the different sampling points. It is possible that the six-point scale is having a knock-on effect on answers to the open-ended expenditure questions in other domains, or, by chance, people asked the six-point scale have higher incomes. The lack of significance of this result following the correction leads us to side with the latter possibility.

respondents find it difficult to recall expenditure. Therefore irrelevant features of the question are employed by the respondent to determine their answer. This is evidenced in our study by a strong effect of the brackets used on the reporting of alcohol expenditures, a substantial effect of time-unit employed on reporting of food and alcohol expenditures and a strong effect of disaggregation in increasing the amount of expenditure reported. It is striking, for example, how few respondents spent more than 2000 euro per year on alcohol compared to the number who spend 40 euro or more per week.

Secondly, expenditure constructs can be difficult to delineate and to understand, often requiring the respondent to adopt a particular interpretation of the question being asked of them. The failure of over one quarter of responses to make a clear distinction between "you" and "your household", and the fact that this interpretation varies across domains is particularly striking. This bias is not limited to expenditure and may also have strong effects on elicitation of assets, bequests and so on. Much further work is necessary to develop protocols to minimise this bias in large population surveys.

Our results raise a number of new questions. Of prime importance is the psychological mechanism underlying these survey effects. Why do respondents, for example, respond differently when asked to report expenditure on a weekly as opposed to yearly timescale? Understanding these mechanisms will facilitate the development of more accurate self-report measures. Despite the instability in reported expenditure across experimental conditions, our results already demonstrate the potential use of prompts and clarifications as methods of alleviating survey bias. The advent of web surveying facilitates the development of expenditure questions that are appropriate to the domain and respondent. For example, the development of interactive web surveys allows questions to be clarified if necessary, without imposing an unnecessary burden on respondents (Schober, Conrad and Fricker, 1999). The results of our paper demonstrate the importance of understanding the effects of question wording when using economic survey data. Our results also provide strong avenues for future research to understand and rectify potential biases.

6. References

Ariely, D., Loewenstein, G., and Prelec, D. (2003), 'Coherent Arbitrariness: stable demand curves without stable preferences', *Quarterly Journal of Economics*, vol. 118, pp. 73-105.

Browning, M., Crossley, T.F., & Weber, G. (2003), 'Asking consumption questions in general purpose surveys', *The Economic Journal*, vol. 491, pp. 540-567.

Conrad, F. G., Schober, M. F., & Coiner, T. (2007), 'Bringing features of human dialogue to web surveys', *Applied Cognitive Psychology*, vol. 21(2), pp.165-187.

Dawes, J., Riebe, E., & Giannopoulos, A. (2000), 'The impact of different scale anchors on responses to the verbal probability scale', *ANZMAC 2000: Visionary Marketing for the 21st Century: Facing the Challenge.*

Delaney, L. and O'Toole, F. (2006), 'Willingness to pay: individual or household?', *Journal of Cultural Economics*, vol. 30, pp. 305-309.

Delaney, L., and O'Toole, F., (2008). "Individual, household and gender preferences for social transfers," *Journal of Economic Psychology*, Elsevier, vol. 29(3), pages 348-359, June.

Diamond, P. A. and Hausman, J.A. (1994), 'Contingent Valuation: Is some number better than no number?', *Journal of Economic Perspectives*, vol. 8, pp. 45-45.

Garland, R. (1991), 'The midpoint on a rating scale: Is it desirable?', *Marketing Bulletin*, vol.2, pp.66-70, Research note 3.

Haisley, E., Mostafa, R., Lowenstein, G. (2008), 'Myopic risk-seeking: The impact of narrow decision bracketing on lottery play', *Journal of Risk and Uncertainty*, vol.37 (1), pp. 57-75.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics6, 65-70.

Hurd, M. (1999), 'Anchoring and acquiescence bias in measuring assets in household surveys,' Journal of Risk and Uncertainty, vol. 19 (1-3), pp. 111-136.

Kahneman, D., Ritov, I., Jacowitz, K, & Grant, P. (1993), 'Stated willingness to pay for public goods: a psychological perspective', *Psychological Science*, vol. 4(5), pp. 310-315.

Kahneman, D., Ritov, I., & Schkade' D. (1999), 'Economic preferences or attitude expressions?: an analysis of dollar responses to public issues', *Journal of Risk and Uncertainty*, Springer. vol. 19, 203-235.

Lindhjem, H., and Navrud, S. (2008), 'Asking for individual or household willingness to pay for environmental goods? Implication for aggregate welfare measures', Munich Personal RePEc Archive.

Loomis, J., Lockwood, M., and DeLacy, T. (1993), 'Some empirical evidence on embedding effects in contingent valuation of forest protection', *Journal of Environmental Economics and Management, vol.* 25(1), pp. 45-55.

Menon, G. (1993), 'The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies', *Journal of Consumer Research, vol.* 20(3), pp. 431-440.

Menon, G., Raghubir, P. and Schwarz, N. (1995), 'Behavioral Frequency Judgments: An Accessibility-Diagnosticity Framework', *Journal of Consumer Research*, University of Chicago Press. vol. 22, 212-228.

Menon, G. and Yorkston, E. A (2000), 'The use of memory and contextual cues in the formation of behavioral frequency judgments', T*he Science of Self-report: Implications for Research and Practice,* Lawrence Erlbaum, pp. 63–79.

Morewedge, C. K., Gilbert, D. T., Keysar, B., Berkovits, M., and Wilson, T (2007), 'Mispredicting the hedonic benefits of segregated gains', *Journal of Experimental Psychology General*, vol. 136**: 700.**

Oswald, A. J. 2008, 'On the curvature of the reporting function from objective reality to subjective feelings', *Economics Letters* 100:369–72.

Pradham, M., (2009). Welfare Analysis with a Proxy Consumption Measure Evidence from a repeated experiment in Indonesia, Fiscal Studies.

Prelec, D., & Loewenstein, G. (1991), 'Decision making over time and under uncertainty: a common approach', *Management Science*, vol. 37, pp. 770-786.

Raghubir, P. and G. Menon (2005), 'When and why is ease of retrieval informative?', *Memory and Cognition***,** vol. 33(5), 821-832.

Schkade, D. A. and Payne, J.W. (1994), 'How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation', *Journal of Environmental Economics and Management*. vol. 26**,** pp. 88-109.

Schober, M. F., Conrad, F. G., & Fricker, S. (1999), 'When and how should survey interviewers clarify question meaning?', *Proceedings of the American Statistical Association, Section on Survey Methods Research*, pp. 985-990.

Schuman, H. and S. Presser (1996), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* London: Sage.

Weber, B. J. and Chapman, G. B (2005), 'Playing for peanuts: Why is risk seeking more common for low-stakes gambles?', *Organizational Behavior and Human Decision Processes*, vol. 97(1), pp. 31-46.

Winter, J. (2002), Bracketing effects in categorized survey questions and the measurement of economic quantities. Sonderforschungsbereich 504 discussion paper. Retrieved, 21st July 2009, from:
http://www.sfb504.uni-mannheim.de/publications/dp02-35.pdf

Winter, J. (2004), 'Response bias in survey-based measure of household consumption', *Economic Bulletin*. **3:** 1-12.

Wirtz, D., Kruger, J., Scollon, C., and Diener, E. (2003), 'What to do on spring break? The role of predicted, on-line, and remembered experience in future choice,' *Psychological Science*, vol. 14(5), pp. 520-524.

Zaller, J. (1992), *The Nature and Origins of Mass Opinion*, Cambridge: Cambridge University Press.

Zaller, J. and Feldman, S. (1992), A simple theory of the survey response: answering questions versus revealing preferences. *American Journal of Political Science*, vol. 36, pp. 579-616.

Table 1: Predictors of Randomisation Conditions

VARIABLES	(1) stopthink	(2) prompt	(3) sixscale	(4) timescale	(7) household	(8) Christmas
Age	0.026*	-0.007	0.010	-0.021	0.021	-0.020
	(0.015)	(0.015)	(0.015)	(0.033)	(0.015)	(0.015)
Female	-0.032	0.009	0.005	0.059	-0.029	-0.002
	(0.032)	(0.032)	(0.032)	(0.071)	(0.032)	(0.032)
Cohabit	0.031**	0.003	0.002	0.024	0.030**	-0.066***
	(0.014)	(0.014)	(0.014)	(0.031)	(0.014)	(0.014)
Kids	-0.002	0.015	-0.020	0.032	-0.011	0.021
	(0.016)	(0.016)	(0.016)	(0.037)	(0.016)	(0.017)
Kidsathome	-0.008	-0.013	0.009	-0.061*	-0.004	-0.014
	(0.014)	(0.014)	(0.014)	(0.032)	(0.014)	(0.014)
Carduse	0.004	-0.014	-0.004	-0.027	0.018	-0.016
	(0.013)	(0.013)	(0.013)	(0.028)	(0.013)	(0.013)
Mortgage	0.049	-0.000	0.021	0.029	-0.022	0.000
	(0.042)	(0.042)	(0.042)	(0.095)	(0.042)	(0.042)
Car	-0.074**	-0.029	-0.003	-0.048	-0.038	0.003
	(0.036)	(0.036)	(0.036)	(0.082)	(0.036)	(0.037)
Missingfinance	-0.005	0.034	-0.029	-0.037	0.005	0.064*
	(0.037)	(0.037)	(0.037)	(0.084)	(0.037)	(0.037)
Missingkids	0.090	-0.076	0.026	-0.058	-0.022	0.262***
	(0.058)	(0.058)	(0.059)	(0.131)	(0.059)	(0.051)
Observations	1042	1042	1042	1042	1042	1042

Notes:
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Predictors of Annual Expenditure

VARIABLES	(1) Food	(2) Total	(3) Alcohol
Stopthink	-0.001	-0.066	0.010
	(0.049)	(0.065)	(0.026)
Sixscale	0.118**	0.165**	0.074***$^\psi$
	(0.049)	(0.065)	(0.026)
Household	0.288***$^\psi$	0.317***$^\psi$	0.053**
	(0.049)	(0.065)	(0.026)
Monthly	-0.196***$^\psi$	-0.044	-0.033
	(0.060)	(0.078)	(0.030)
Annual	-0.579***$^\psi$	-0.224***$^\psi$	-0.154***$^\psi$
	(0.059)	(0.079)	(0.028)
Christmas	-0.045	-0.025	-0.067***$^\psi$
	(0.049)	(0.065)	(0.026)
Constant	7.114***	8.975***	
	(0.062)	(0.081)	
Observations	1044	996	1142
R-squared	0.118	0.040	.

Notes:

Holm – Bonferroni correction applied (Holm 1979) with 2 assumed true null hypotheses from 20 total hypotheses. $^\psi$ indicates significant at the 5 per cent level following correction. For Food and Total, coefficients represent the box-cox adjusted OLS estimates. For Alcohol, coefficients represent marginal effects from a probit model of probability of consuming greater than 40 euro per week or equivalent. Food Expenditures greater than 50,000 per year and Total Expenditures greater than 80,000 per year are removed as outliers.

Table 3: Effects of Individual/Household Clarification

	Domain	Respondents' interpretation:		
		Individual	Mixture	Household
Question wording: "You"	Alcohol	69 %	11 %	20 %
	Food	51 %	16 %	33 %
	Total	74 %	6 %	19 %
Question wording: "Your household"	Alcohol	24 %	28 %	48 %
	Food	10 %	25 %	65 %
	Total	18 %	16 %	66 %

Notes:
Table is limited to respondents who have more than one individual (including themselves) in their household.

Table 4: Summary of Experiments and Hypotheses

	Hypothesis	Procedure	Results
Question Interpretation and Human Dialogue: You and Your Household	Survey responses are not sensitive to the distinction between individual expenditure and collective household expenditure.	Respondents were randomly assigned to report either how much "you" spent or how much "your household" spent on food, alcohol and in total. A follow-up question asked whether responses referred to their individual expenditure; their household expenditure; or a combination of the two.	Respondents struggled to differentiate between their individual expenditure and that of their household. One in every five respondents interpreted "you" as referring to their household when estimating their expenditure on alcohol and their total expenditure. One third of the sample reported their household expenditure when asked to report "your" expenditure on food. Around two thirds of respondents living with a partner or relatives reported their household expenditure on food when asked to do so.
Human Dialogue 1: Stop and Think	Respondents primed with a stop-and-think prompt will report a higher total expenditure than others because they access a wider range of considerations.	Prior to answering questions on expenditure, a random subsample received a stop-and-think prompt.	The stop-and-think prompt had no effect on reported expenditure amounts.
Constructed Responses 1: Disaggregated Prompts	An itemised list of disaggregated motoring expenses will help respondents recall motoring expenditure that would otherwise be forgotten.	A random subsample of respondents received an itemised list of motoring expenses e.g. fuel; insurance; vehicle maintenance etc.	Respondents in the disaggregated condition reported significantly higher levels of expenditure (n = 192; t= 2.79; p = 0.006).
Constructed Responses 2: Timescale Effects	People report lower expenditure per day as the timescale increases.	Respondents were randomly assigned to report their expenditure on a weekly, monthly or yearly timescale.	The effect of timescale is highly significant in the hypothesised direction.
Constructed Responses 3: Anchoring to brackets	Reported expenditure is sensitive to the bracketed categories chosen by the survey designer.	Respondents were randomly assigned to report their alcohol expenditure on either a higher or lower anchored scale.	The probability of reporting having spent over €40 or equivalent is higher if alcohol expenditure is elicited on the higher anchored scale (z = 2.83; p = 0.007).
Constructed Responses 4: Recency bias	Respondents surveyed one week before Christmas will report having spent more on alcohol and food in a typical week over the past year than will respondents surveyed after Christmas.	Half of the sample answered the survey one week before Christmas. The other half was recruited in mid-January.	Christmas has no effect on food expenditure estimates or aggregate total expenditure estimates and reduces the likelihood of reporting high average alcohol expenditure.