How Successful I am Depends on What Number I Get: The Effects of Numerical Scale Labels and Need for Cognition on Survey Responses

Ting Yan
Survey Research Center, University of Michigan

Acknowledgment. The work reported here was part of my Ph.D. dissertation submitted to the Graduate School, University of Maryland. I am especially grateful to my academic advisor Roger Tourangeau and my committee members for their help and advice. I also like to thank Norbert Schwarz and Frauke Kreuter for their comments on an earlier draft.

Abstract

To further the understanding on the mechanism of numerical scale values, this study varied the numerical labels for scale points and examined its effect in relation to individual respondent's need for cognition. Embedded in a web survey, this experiment succeeded in producing a variety of evidence that respondents worked out an inference from the numerical values and based their responses on that inference. The shift in responses induced by the numerical scale values was unexpectedly robust; when a scale started with a negative number, it pushed the responses to the right or the positive end of the scale across items and across fonts. Process measures (such as recall) and answers to the retrospective probe confirmed that respondents paid attention to the numerical labels on the scales and used them to interpret the verbal labels on the end points of the scale.

However, not everyone was affected by the numbers on the scales. The hypothesized effect of the negative scale values was observed only among respondents with a high need for cognition, but not among those with a low need for cognition. This finding seemed to suggest two things. First, Gricean effects of this sort involve controlled processes; people need to process deeply for the numbers to affect the answers. Second, unlike the response errors committed by satisficing respondents (who skip or slack off on certain cognitive steps), Gricean effects are an optimizing error or high effort error, committed by respondents who try to be good, cooperative, and thorough respondents.

Keywords: Gricean effect, pragmatic inference, optimizing, satisficing, measurement error, web surveys

Introduction

Prior studies have established that many incidental features of survey questions can affect respondents' answers, creating unwanted response errors (Schwarz 1999; 1996). One line of research, led by Schwarz and his colleagues, attributes this type of measurement error to respondents' use of Grice's Cooperative Principle (CP) and its associated maxims in the survey research setting as in everyday conversation (see Schwarz 1996; for a review, see Tourangeau, Rips, and Rasinski 2000); the term "Gricean effect" is used throughout this paper to refer to this type of measurement error. According to Grice, participants in communication are cooperative and rational; they speak in a truthful, informative, relevant, and clear manner (Grice 1989). Relying on this cooperative principle, respondents make use of various visual features of survey questions in interpreting questions and forming responses; they see the visual features as task elements that are essential to the question-answer process (see Couper, Tourangeau, and Kenyon 2004, for the distinction between style elements and task elements).

The numerical values assigned to a rating scale are one feature that seems to be taken as a task element in the survey response process. Schwarz, Knauper, Hippler, Noelle-Neumann, and Clark (1991) first demonstrated that respondents came to different interpretations of the verbal end points of a scale when the scale ran from 0 to 10 than when it ran from –5 to 5. Presuming that every piece of information was relevant based on the Gicean Cooperative Principle, respondents inferred that the same end label ("not at all successful") meant the "mere absence of noteworthy success" when 0 was assigned to that scale point, but "the presence of failure" when –5 was assigned to that point (Schwarz et al. 1991). As a result, respondents were less likely to select values less than or equal to the midpoint with -5 to 5 scale labels than with the 0 to 10 labels; thus, responses were pushed to the right or positive side of the scale.

In addition to rating scales, Schwarz and his colleagues also investigated the effects of different numerical values attached to a frequency scale (Schwarz, Grayson, and Knauper 1998, Experiment 1). The frequency scale ranged either from 0 to 10 or from 1 to 11. The end labels remained "rarely" for 0 or 1 and "often" for 10 or 11. Again, the numerical values of the scales influenced the responses – respondents reported higher frequencies when the scale ranged from 0 to 10 than when it ranged from 1 to 11. Schwarz and colleagues speculated that the end label "rarely" indicated a lower frequency when combined with value 0 than with value 1; as a result, the scale running from 0

to 10 shifted the means to the higher end of the frequency (Schwarz et al. 1998).

This effect of numerical labels is replicated on questions about different topics (Haddock and Carrick 1999). The numerical values assigned to scale points were shown to consistently affect survey responses (regardless of the types of the scales used). The effect is robust across modes of administration (telephone interview vs. self-administered questionnaires vs. face-to-face interviews), for both unipolar and bipolar scales, for various domains, and for self and proxy reports (O'Muircheartaigh et al. 1995; Schwarz et al. 1991; Schwarz and Hippler 1995). Nonetheless, two important issues have yet to be addressed to gain a fuller understanding of the mechanism of numerical scale values.

The first one is the lack of direct evidence that the response changes induced by the numerical values were due to respondents' utilization of the maxim of relation during the survey response process. In most of the past studies, changes in response distributions were observed and were speculated to be caused by respondents' interpretation of scale labels as a result of their utilization of the maxim of relation. Only one experiment out of the eight documented sought direct evidence of respondents' interpretation of the scale labels through a follow-up question.

Second, the existing studies mainly looked at the overall main effects of the numerical labels, implicitly assuming that this particular type of Gricean effect triggered by the numerical labels of the scales is across-the-board among all respondents. However, respondents differ in how they answer survey questions and how much cognitive effort they exert in survey response process. If taking numerical labels into consideration when forming responses involved additional cognitive effort (i.e., attention to and processing of the numbers), there would be some respondents who were either unable to or unwilling to expend the additional cognitive effort in processing the numerical scale values. Therefore, existing studies might have covered up subgroup differences with regard to the effects of numerical labels by focusing only on main effects.

Two closely related notions have to do with individual differences in exerting cognitive efforts. One concept is people's need for cognition, which represents "the extent to which people tend to engage in and enjoy effortful cognitive activity" (Petty and Jarvis 1996, p.221). According to Petty and Jarvis, people differ in the cognitive effort they exert in thought processes. Those with a high need for cognition (HNC) tend to seek out more information and process information more carefully whereas the group with a low need for cognition (LNC) is more inclined to adopt simple or cognitively untaxing strategies before making an evaluation (Petty and Jarvis 1996; see Cacioppo, Petty, Feinstein, and Jarvis 1996 for a comprehensive review on need for cognition). The need for cognition can be considered a person's tendency to think and enjoy the process of thinking; thus, it reflects a stable personality trait rather than a temporary choice of cognitive strategy. It is typically assessed with an 18-item need for cognition scale created by Cacioppo and colleagues (Cacioppo, Petty, and Kao 1984).

In the survey context, Petty and Jarvis (1996) predicted that the HNC group is expected to process survey questions more carefully than the LNC group. Accordingly, they speculated that the HNC group is more susceptible to high effort biases such as primacy effects and priming effects while low effort biases – e.g., errors resulting from yea-saying to agree-disagree items (or acquiescence) or providing similar responses to a batch of questions (non-differentiation) – are produced by the LNC group only (Petty and Jarvis 1996).

Krosnick's notion of satisficing distinguishes two types of respondents based on their response styles: "optimizers" respond carefully and thoughtfully whereas "satisficers" short cut their cognitive processes by either executing cognitive steps less completely or skipping certain cognitive steps (Krosnick 1991; 1999). Consequently, satisficers are more attracted to cognitively untaxing response behaviors such as giving "Don't Know" responses, acquiescing answers, non-differentiated responses, and picking the first seemingly reasonable response option (Krosnick 1991; 1999; Narayan and Krosnick 1996).

Unlike the need for cognition, satisficing reflects a respondent's response strategy and the likelihood of satisficing is a function of respondents' cognitive ability, motivation, and task difficulty (Krosnick 1991, 1999). There is not yet one method to assess satisifcing but Krosnick (1991, 1999) provided a list of covariates that are said to be associated with satisficing. For instance, the need for cognition affects the likelihood of satisficing through their impact on respondent cognitive ability whereas fatigue and boredom show their influence through motivation. Type and structure of survey questions have to do with task difficulty.

Though different in conceptualization and operationalization, both the need for cognition and satisficing predict individual differences in the amount of cognitive effort exerted in survey response process. Both predict that those respondents who exert less cognitive effort respond to the scale numerical values differently from those who exert more effort.

Despite the large number of research studies on need for cognition reviewed in Cacioppo et al. (1996), need of cognition is not commonly considered in the survey context. There are only a handful of studies looking at the effect of need of cognition on survey

responses. Bizer, Krosnick, Holbrook, Petty, Rucker, and Wheeler examined the moderating effect of need for cognition in the 1998 National Election Survey Pilot Study (2002). They found out that people low in need for cognition were less likely to enjoy the survey process and were slightly more likely to say "don't know" when asked attitude questions (Bizer et al. 2002). However, Fournier, Lyle, Cutler, and Soroka, using the same dataset, failed to confirm the hypothesis that need for cognition is related to susceptibility to opinion change (2004). One reason for the weak effect of need for cognition in the two studies lies in the inadequate measure of need for cognition; only two items out of the original 18-item scale were included in the pilot study.

McCabe and Brannon (2004) reported a partial replication of the general-specific questions with an emphasis on the impact of a joint lead-in on responses. Both Schwarz, Strack, and Mai (1991) and Tourangeau, Rasinski, and Bradburn (1991) demonstrated that, when a specific question (e.g., relationship satisfaction) preceded a general question (e.g., satisfaction with life in general), the correlation between the two questions was reduced in the presence of a joint lead-in because respondents applied the maxim of quantity to avoid providing redundant information. McCabe and Brannon found that only the HNC respondents displayed such an attenuated correlation between the items, but not the LNC respondents. Their finding suggested that the conversational norm to avoid redundancy was not automatically applied in the survey context; only those with a high need for cognition seemed to apply the maxim. This is the first empirical evidence for the prediction that the HNC group is subject to high effort bias – errors resulting from respondents' optimizing behavior.

This study aimed to fill the gap in existing literature, fulfilling two goals: a) seeking direct evidence that respondents use the maxim of relation when answering rating scales questions with numerical labels, and b) demonstrating that the use of the maxim of relation in this case is a high effort bias committed only by respondents with a high need for cognition (HNC).

Three process measures and one follow-up question are employed to provide direct evidence of respondents' utilization of the maxim of relation. The corresponding hypotheses regarding the process measures are that a negative scale number will induce better recall, greater attention and usefulness rating by respondents. The follow-up question solicits directly respondents' inference about the scale label; respondents are hypothesized to infer a presence of negative trait when the scale starts with a negative number. In addition, given the differential cognitive effort exerted in the question-answer process, respondents high in need for cognition are hypothesized to utilize the maxim of relation, showing larger effects of negative scale numbers than people low in need for cognition will.

The Study

Overview. This experiment was embedded in a web survey conducted by MS Interactive. Survey Sampling Inc. (SSI) selected the sample for this study from its opt-in Web panel (Survey Spot) of over one million persons who have signed up online to receive survey invitations. SSI selected 17,362 e-mail addresses for this study and sent out e-mail messages inviting the recipients to take part in "a study of attitudes and lifestyles." The e-mail invitations included the web address (URL) for the survey web site and a unique identification number (which prevented respondents from completing the survey more than once). The survey ran from May 24 to June 2, 2005. Of the 17,362 invited to participate in the survey, 1,071 completed the entire survey (and 146 others got part way through) for a response rate (AAPOR [2000] RR1) of 6%. The questionnaire included questions on a range of topics, most of them attitudinal. The 18-item need for cognition scale (Cacioppo, Petty, and Kao, 1984) was included in the last section, together with demographic questions. This experiment came first in the questionnaire.

Experimental manipulation. This experiment manipulated both the numerical values assigned to the scale points (replicating the earlier studies) and the appearance of the scale values in a 2 (numerical labels: 0 to 6 vs. -3 to 3) x 2 (appearance: normal font vs. faint font) factorial design.[1] Table 1 displays the number of completes per experimental condition.

Target questions. The key target question is the success item. I used the same question wording as in Schwarz et al. (1991). For replication purpose, respondents were also asked to rate their moodiness, their nervousness, and optimism along one of the four randomly assigned scales. Respondents got the same numerical labels for all four questions.

[1]The faint font version of the scales displayed the numerical values assigned to scale points in a distinct font that was much fainter than the font used for the question text and the verbal label. Such fonts are typically used in paper questionnaires for information that is *not* intended for the respondents. The purpose of this font manipulation is to test whether a faint font would lead respondents to discount the relevance of the scale labels in a web survey. The analysis of the font manipulation is not presented in this paper.

Follow-up questions. The follow-up questions asked respondents about their use of the scale values and the inferences they drew about the scale end labels.[2]

Results

I begin by presenting the analyses of responses to the four target questions, followed by analyses on respondents' use of and inferences about the scale.

Responses. For all four scale conditions, I coded the responses from 1 to 7, where 1 corresponded either to 0 or -3, and 7 to 6 or 3. To compare responses to scales with different numerical values, I examined the mean ratings of the 0 to 6 scales and of the -3 to 3 condition (see Table 2).

As evident from Table 2, the negative scale values produced a mean shift in the key target item (success), replicating the finding by Schwarz et al. (1991). In addition, mean shifts were replicated on the other three items too. One-way ANOVAs, conducted on all four target questions, confirmed that numerical values had a significant effect on responses for the first three items, but not on the last one.

To examine the effects of an individual's need for cognition on responses, I split the sample into two groups based on their scores on the need for cognition scale. Those who scored higher than 3.5 (median value) are considered to have a high need for cognition (HNCs) whereas those with a score of 3.5 or lower are regarded as the LNC group. Figure 1 plots the mean responses to the success item by the numerical labels and the need for cognition. The negative scale number induced a mean shift for both groups; however, consistent to my hypothesis, the shift was significantly bigger for respondents with a high need for cognition than for those with a low need for cognition.[3] The main effects of numerical labels ($F(1,1045)=13.46$, $p<.00$), of need for cognition ($F(1,1045)=18.57$, $p<.00$), and the interaction effect ($F(1,1045)=4.79$, $p<.05$) are highly significant.

Process measures. I assessed the inferences respondents drew based on their answers to follow-up questions. I first examined various process measures. If conversational implicatures were worked out, the extra effort needed to work out an implicature should produce better recall of the numbers that triggered the interpretative maxim. Respondents should have also paid more attention to the numbers and considered the numerical values more useful when they used them in interpreting the response scale.

I examined the percentage of respondents who recalled the leftmost scale value correctly by the scale values and need for cognition. Figure 2 indicates that, in general, more respondents recalled the number correctly when presented with the -3 to 3 scale labels than with the 0 to 6 labels ($\chi^2=3.18$, $p=0.07$). However, the HNC respondents were more likely to recall correctly the number assigned to the leftmost scale point than the LNCs; the difference between the two respondent groups was bigger when the scale started with -3 rather than 0 (see Figure 2). The difference in the percentage of correct recall by numerical label conditions is marginally significant for the HNC group ($\chi^2=3.53$, p=.06), but not for the LNC group ($\chi^2=0.56$, ns). Still, the pattern is consistent with the prediction – the negative values were recalled better overall and the difference in recall is more marked with the HNC group.

Another two follow-up items asked respondents how much attention they paid to the numerical values attached to the scale and how useful they considered those numbers. Figure 3 plots the average ratings of attention and usefulness for the numerical labels and the need for cognition, demonstrating that respondents tended to pay more attention and consider the scale label more useful when the scale started with a negative number than with zero. The effect of numerical values is significant for the attention ratings ($F(1,1066)=15.73$, $p<.0001$), but not significant for the usefulness ratings.

Compared to the LNC group, the HNC respondents consistently claimed to have paid more attention to the negative numerical label (simple main effect of numerical label: $F(1,1066)=13.37$, $p<.000$) and to have considered the negative number more usefully ($F(1,1059)=3.15$, $p<.10$) when the scale ran from -3 to 3. The differences in ratings between the two numerical label conditions are more perceptible with the HNC respondents than with the LNC group; the relevant simple main effects of numerical labels are not significant for the LNC respondents.

Inferences. Process measures such as recall task, and self-reported attention level and usefulness of numbers suggested that inferences were drawn in response process. In order to seek direct evidence and to determine the exact inference respondents drew from the numerical values, the last follow-up question asked respondents what the scale label "not at all successful"

[2] The exact wordings of the target questions and follow-up questions are not presented, but can be requested from the author.

[3] Analyses of the other three items revealed a significant three-way interaction (item*numerical label*need for cognition). One post-hoc explanation was the confounding of the connotative meanings of the moodiness and nervousness items with need for cognition. The HNC respondents were less likely to consider themselves as nervous or moody; as a result, they tended to place them on the left side of the scale, rather than the right side of the scale.

meant to them.[4] There were six answer categories to this question (see bottom panel of Appendix for the exact wordings of the six answer categories). I collapsed the answer categories into two groups – one group represents the absence of success and the other group the presence of failure. Figure 4 plots the percentage of people inferring "presence of failure" by scale numerical labels and the need for cognition.

The result supports the conjecture of Schwarz and his colleagues (1991) about how respondents interpret the scales with different numerical labels. Significantly more respondents interpreted the scale label "not at all successful" to mean the presence of failure when the numerical labels ran from -3 to 3 than when they ran from 0 to 6 (χ^2=4.80, p=.03), suggesting that respondents did take the numerical values into consideration when they constructed their answers.

However, it is again the HNC group who used the numerical values in interpreting scale verbal labels; significantly more HNC respondents drew the inference of "presence of failure" when the scale started with -3 (73%) than when the scale ran from 0 to 6 (63%) (χ^2=4.97, p=.03). By contrast, about same percentage of LNC respondents interpreted the scale label "not at all successful" to mean the presence of failure when the numerical labels ran from -3 to 3 (62%) than when they ran from 0 to 6 (65%); the difference is not significant controlling for the need for cognition (χ^2=.88, ns). Figure 4 demonstrates that it is respondents with a high need for cognition that drew inferences from the negative scale number, replicating the finding by McCabe and Brannon (2004) and supporting the claim by Petty and Jarvis (1996) that the HNC respondents are more susceptible to high effort biases.

Discussion

Several studies by Schwarz and his colleagues demonstrated that the numerical values assigned to the scale points affect the distribution of the responses. This study replicates and extends previous work by Schwarz and colleagues by manipulating the numerical labels for scale points and examining the moderating effect of need for cognition. The results showed that the mean shift in response to the right side of the rating scale induced by negative numerical labels was robust across items.

This study also provides direct evidence that respondents draw inferences about the verbal labels of the scale points based on the Gricean maxim of relation. Process measures such as recall task, self-reported attention to the scale numbers, and the retrospective probe confirmed that respondents paid attention to the numerical labels on the scales, carefully processed the negative numbers, and worked out inferences to interpret the verbal labels on the end points of the scale. Process measures also seemed to suggest that the processing of negative numerical labels is a controlled process (evidenced by better recall).

This study further showed that this Gricean effect is a high-effort bias committed by optimizing respondents; the effect of negative scale numbers is more marked with respondents who have a high need for cognition than those with a low need for cognition. In other words, more thinking and deeper processing by respondents high in need for cognition made the effects of negative scale numbers bigger. This finding, together with the results from process measures, showed that processing negative scale values is a controlled process. Consistent with the finding by McCabe and Brannon, this study provides further empirical evidence that the resulting Gricean effect of negative scale values is a high effort bias committed by respondents who read too much into survey questions and contexts.

This study points to a few issues that merit survey researchers' attention. First, research on survey measurement errors have been focusing on identifying and fixing response errors committed by respondents who are cognitive misers and who haven't put in as many cognitive effort as we desire (Krosnick 1991; 1999). The implicit assumption held by most survey researchers is that more thinking and deeper processing is better than little thinking and shallow processing. However, whether this is true or not may depend on the specific response effect. As this study showed, careful processing could lead to errors as much as the lack of careful processing. Therefore, survey researchers should shift away from their traditional emphasis on errors committed by satisficing respondents to those by optimizing respondents.

Second, survey researchers should start develop techniques for overcoming optimizing errors or high-effort biases. Traditional techniques for overcoming errors resulting from satisficing – such as motivating respondents – might not effectively reduce errors by optimizing respondents. At least, encouraging respondents to think more carefully about the question and the negative scale numbers would only trigger the controlled processes and increase the mean shift caused by the negative scale numbers.

Third, Schwarz suggested that the best way to reduce Gricen effect is for survey researchers to become a cooperative communicator (Schwarz 1998; 2000). Survey researchers have the responsibility to communicate to respondents what should be perceived as informative and what should not be. Thus, it is no longer enough to simply pretest survey questionnaires at the level of semantics and syntax; survey researchers

[4] I present here the results based on the closed-end question. Analyses of the open-ended responses were similar and didn't change the conclusions reported here.

should also pretest survey instruments at the level of pragmatics in order to prevent Gricean effects reported here from happening (Tourangeau, Rips, and Rasinski 2000).

Last, given that people with different levels of need for cognition are subject to different types of measurement error, survey researchers should consider including the need for cognition scale in their instrument; the inclusion of such a scale could shed light on respondents cognitive processing and efforts. It could also be used as a covariate to be controlled for in their statistical modeling and analysis.

References

Bizer, George, Jon Krosnick, Allyson Holbrook, Richard Petty, Derek Rucker & Christian Wheeler. 2002. "The impact of personality on political beliefs, attitudes, and behavior: Need for cognition and need to evaluate", Paper presented at the annual meeting of the American Political Science Association.

Cacioppo, John, Petty, R. E., Feinstein, J., & Blair Jarvis. 1996. "Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition", *Psychological Bulletin*, 119: 197-253.

Cacioppo, John T., Richard E. Petty, and Chuan Feng Kao. 1984. "The Efficient Assessment of Need for Cognition." *Journal of Personality Assessment* 48:306-07.

Couper, Michael. P., Roger Tourangeau, and Kristen Kenyon. 2004. "Picture this! Exploring Visual Effects in Web Surveys." *Public Opinion Quarterly* 68:255-66.

Fournier, Patrick, Greg Lyle, Fred Cutler, and Stuart Soroka. 2004. "Need for Cognition, Need to Evaluate, and Electoral Attitude Change," Paper presented at the annual meeting of the American Association for Public Opinion Research.

Grice, H. Paul. 1989. *Studies in the Way of Words*. Harvard University Press.

Groves, Robert M. 1989. *Survey Eerror and Survey Cost.* New York: John Wiley and Sons.

Haddock, Geoffrey, and Rachael Carrick. 1999. "How to Make a Politician More Likeable and Effective: Framing Political Judgment Through the Numeric Values of a Rating Scale." *Social Cognition* 17:298-311.

Krosnick, Jon A. 1991. "Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-36.

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

McCabe, Amy E. and Laura A. Brannon. 2004. Application of Conversational Norms to the Interpretation of Survey Results as a Function of Participants' Need for Cognition." *The Journal of Psychology* 138:91-4.

O'Muircheartaigh, Colm A., George D. Gaskell, and Daniel B. Wright. 1995. "Weighing Anchors: Verbal and Numeric Labels for Response Scales." *Journal of Official Statistics* 11:295-307.

Petty, Richard E., and W. Blair G. Jarvis. 1996. "An Individual Differences Perspective on Assessing Cognitive Processes." In *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research,* ed. Norbert Schwarz and Seymour Sudman, pp. 221-58. San Francisco: Jossey-Bass Publishers

Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Erlbaum.

Schwarz, Norbert. 1998. "Warmer and More Social: Recent Developments in Cognitive Psychology." *Annual Review of Sociology* 24:239-64.

Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54: 93-105.

Schwarz, Norbert. 2000. "Social Judgment and Attitudes: Warmer, More Social, and Less Conscious." *European Journal of Social Psychology* 30: 149-76.

Schwarz, Norbert, Grayson, Carla E., and Barbel Knauper. 1998. "Formal Features of Rating Scales and the Interpretation of Question meaning." *International Journal of Public Opinion Research* 10:177-84.

Schwarz, Norber, and Hans-Jurgen Hippler. 1995. "The Numeric Values of Rating Scales: A Comparison of Their Impact in Mail Surveys and Telephone Surveys." *International Journal of Public Opinion Research* 7:72-4.

Schwarz, Norbert, Barbel Knauper, Hans-Jurgen Hippler, Elizabeth Noelle-Neumann, & Clark, Leslie. (1991). "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55: 570-82.

Schwarz, Norbert, Fritz Strack, and Hans-Peter Mai. 1991. "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis." *Public Opinion Quarterly* 55: 3-23.

Tourangeau, Roger, Kenneth A. Rasinski , and Norman Bradburn. 1991. "Measuring

Happiness in Surveys: A Test of the Subtraction Hypothesis." *Public Opinion Quarterly* 55: 255-66.

Tourangeau, Roger, Lance Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press.

Table 1. Number of Completes Per Experimental Condition

	0 to 6	-3 to 3	Total
Normal Font	259	271	530
Faint Font	271	270	541
Total	530	542	1071

Table 3. Mean Responses By Numerical Labels and Significance Tests

	0 to 6	-3 to 3	Significance Tests	
Success	4.86	5.13	$F(1,1061)=13.04$	$p<.001$
Moodiness	3.54	3.79	$F(1,1064)=7.13$	$p<.01$
Nervousness	3.25	3.43	$F(1,1061)=3.87$	$P<.05$
Optimism	5.06	5.15	$F(1,1065)=1$	ns

Figure 1. Mean Ratings of Success by Numerical Labels and Need for Cognition

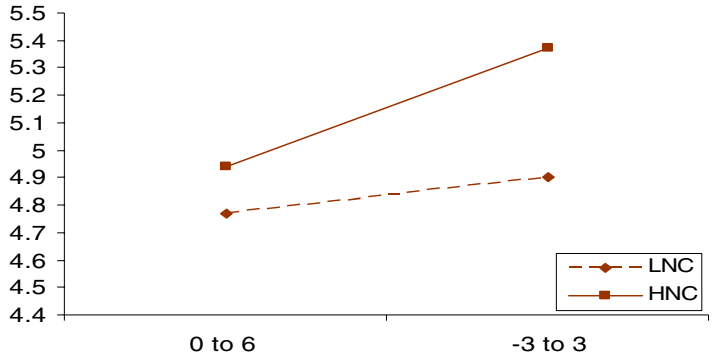

Figure 2. The Percentage of Correct Recall by Numerical Labels and Need for Cognition

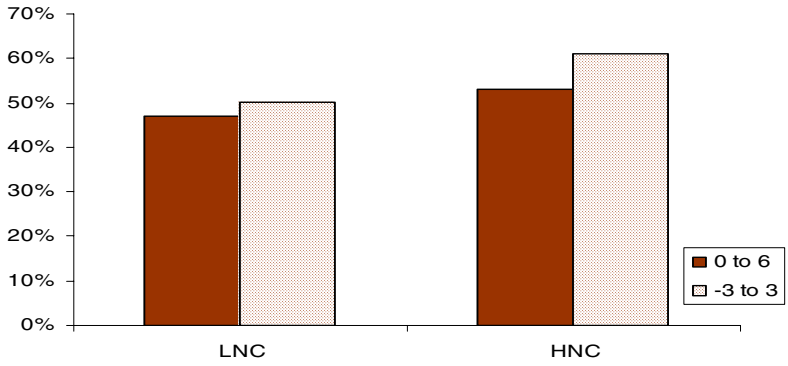

Figure 3. Attention and Usefulness Rating by Numerical Labels and Need for Cognition

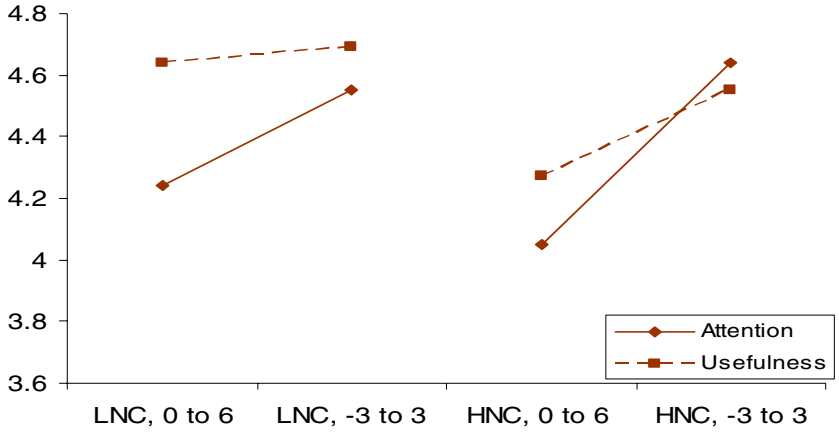

Figure 4. Percentage of Respondents Inferring "Presence of Failure" by Scale Numerical Values and Need for Cognition