It's a Trap! The Influence of Instrumental Manipulation Checks on Response Non-

Differentiation and Gricean Norm Effects

Aashna Sunderrajan

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of Bachelor of Science with

Honors in Biopsychology, Cognition and Neuroscience from

the University of Michigan

2014

Advisors: Dr. Norbert Schwarz and Dr. Allison Earl

Graduate Mentor: David J. Hauser

Authors Note

Aashna Sunderrajan, Department of Psychology, University of Michigan, Ann Arbor.

Correspondence concerning this paper should be addressed to Aashna Sunderrajan, 1770 Broadway Street, Ann Arbor, MI 48105. Email: aashnas@umich.edu.

Abstract

The instructional manipulation check (IMC) is a methodological tool designed to measure whether participants read the instructions in a survey, and thus, it is frequently used in online studies to identify inattentive participants. However, research on survey design has suggested that participants can infer intentions from prior questions, affecting responses to later questions. As the IMC tells participants of a communicative context wherein the experimenter intends to trick participants into giving wrong answers, this study assessed whether the simple inclusion of an IMC could lead to additional cognitive processing that subsequently reduced participants' satisficing behaviors (Study 1), while increasing classic Gricean norm effects (Study 2). The results showed that the IMC did not affect participants' satisficing behaviors or their responses to several of the Gricean norm effect measures, with one exception. Implications and future research areas are discussed.

Keywords: instrumental manipulation checks, survey methods, context effects, cognitive processing

It's a Trap! The Influence of Instrumental Manipulation Checks on Response Non-differentiation

and Gricean Norm Effects

In survey research and online experiments, inattentive participants are a common source of error and context effects. These participants increase noise, reduce statistical power and decrease the validity of the data. To manage this concern, research has examined ways of explaining and empirically identifying such inattentive participants. Recently, Oppenheimer, Meyvis and Davidenko (2009) proposed that inattentive participants were probably less likely to read the instructions of a survey completely. Based on this premise, the researchers devised the Instrumental Manipulation Check (IMC) in an attempt to identify such inattentive participants.

The IMC is a methodological tool designed to measure whether participants read the instructions in a survey, by asking participants to disregard the standard response format for the survey in lieu of a response that confirms that they have read the instructions. In their study, Oppenheimer et al. (2009) gave an IMC within a packet of unrelated questions to participants and identified that a substantial number of individuals failed the IMC when filing out the survey. However, by using the IMC to identify these inattentive participants, it was possible to exclude them from the data analysis, which subsequently increased the power associated with the experiment. Accordingly, the researchers concluded that the inclusion of the IMC was a useful tool in identifying and filtering out inattentive participants, thereby increasing the reliability of the data set (Oppenheimer, Meyvis, & Davidenko, 2009).

However, the IMC is a unique question, in that it requires participants to override their initial response, identified as an error by the survey creator in the question itself, but which intuitively seems correct. Conversely, in order to arrive at the correct response, participants must input an answer which is correct, as deemed by the survey creator, but which violates the

assumption of a cooperative survey participant. This tells participants of a communicative context wherein the experimenter intends to trick participants into giving wrong answers, and thus, can have the ability to affect participants' mindsets on subsequent responses.

Related research on survey design does in fact suggest that minor aspects of surveys can affect people's cognitive states, and can certainly affect survey responses. Questions can render information accessible, making it more likely to be used in later judgments, such as in follow-up questions (Srull & Wyer, 1979). Additionally, participants can infer intentions from the questions and the communicative context of the survey, thereby further influencing their responses to later questions (Hauser & Schwarz, 2013a; Strack, Martin, & Schwarz, 1988).

Hauser and Schwarz (2013a), for example, conducted an experiment to assess whether the simple inclusion of the IMC affected participants' subsequent survey responses. In one study, the researchers administered an IMC (Oppenheimer et al., 2009) to see if it affected responses on the Cognitive Reflection Task (CRT; Frederick, 2005), a measure of an individual's ability to suppress the intuitively driven (incorrect) response that first enters their mind. When the IMC was administered prior to the measure, participants had higher CRT scores. This suggests that the IMC is capable of engaging participants in error detection processing.

Pursuing this further, Hauser and Schwarz (2013b), in an unpublished study, presented participants with the IMC alongside the Moses illusion and Switzerland question (Song & Schwarz, 2008). The Moses illusion is a task of error distortion, with a reliance on the spontaneous answer inhibiting the acknowledgement of the correct answer. The Switzerland question, on the other hand, is an undistorted question, with the spontaneous answer also being the correct answer. Participants presented with the IMC committed less errors on the Moses illusion task, while producing more errors on the Switzerland question. These studies thus

suggest that the IMC signals a need for error detection processing, prompting participants to suppress the intuitively driven response that first enters their mind when encountering certain questions. Hence, the presentation of an IMC in a study seems to be able to increase subsequent cognitive processing on seemingly "tricky" tasks.

As IMCs are often used in research as measures, not manipulations, it is important to study how IMCs affect survey participants' mindsets. Thus, the purpose of this study is to determine whether IMCs function as interventions for common survey context effects, two of which include satisficing behaviors and Gricean norm effects.

Study 1

In their attempt to reduce cognitive demands, participants take mental shortcuts while responding to questions (Krosnick, 1991). Known as satisficing, this behavior produces patterns of inaccurate responding. According to Krosnick (1991), there are six forms of satisficing behaviors often associated with survey work: (a) selecting the first response alternative that seems reasonable, (b) agreeing with the assertions made in the survey, (c) endorsing the status quo rather than social change, (d) failing to differentiate between items on a rating scale, (e) saying "I don't know" rather than stating an opinion and, (f) haphazardly choosing among the response alternatives presented.

Each of these forms of satisficing is problematic to survey research, as they create distortions in participant response patterns. However, strong forms of satisficing (compared to their weaker counterparts) are generally more troublesome to researchers (Krosnick, 1991). Weak forms of satisficing occur when participants are less thorough in discerning the meaning of a question and determining an appropriate response. Strong satisficing, on the other hand, occurs when participants interpret each question on a superficial level and select responses that they

believe to be reasonable, without referring to any internal cues relevant to the question of interest (Krosnick, 1991).

In survey research, these types of satisficing behaviors can be propagated by the survey design. For example, questions concerning similar topics with the same response alternatives are often grouped together (Krosnick, 1991). Although this form of organization might make the survey easier for participants to complete, research has shown that it can also lead to the phenomenon of non-differentiation (Krosnick, 1991; Krosnick & Alwin, 1988). Thus, instead of considering each question individually and attempting to differentiate between the questions in their ratings, participants consider the entire range of questions at once and assign them the same rating, or rate one question then reuse that rating for the others (Krosnick, 1991). This results in a high percentage of questions being rated similarly, which decreases the reliability of the data set and the power associated with it.

In their research, Hauser and Schwarz (2013a) identified that the IMC acts not just as a measure, but as a manipulation that changes the mindset of participants. This suggests that the inclusion of the IMC in survey research may lead to the activation of deeper thought in participants. If IMCs do cause participants to think more while completing a task, while satisficing is a consequence of participants' attempt to reduce cognitive demands, then it follows that the presentation of an IMC before a non-differentiation measure would decrease satisficing behaviors. Therefore, we hypothesize that the presentation of an IMC preceding the non-differentiation question will discourage satisficing, and thus, produce less non-differentiation response patterns.

Method

Participants

Seven hundred and ninety eight participants on Amazon Mechanical Turk (MTurk), a forum service where workers complete short tasks for small monetary rewards, participated in the study (456 male, 342 female). The participants ranged in age from 18 to 81. The survey took between four to six minutes to complete, and each participant was compensated with 40 cents for participating.

Materials and Procedure

Participants were directed to a survey ostensibly on current issues. The IMC and non-differentiation measure were part of a larger battery of surveys that included 8 tasks. Thus, participants completed a series of questions, including an IMC and a non-differentiation measure, but were randomly assigned to complete them in different orders. Half of the participants received the IMC as the first question in the survey packet, whereas the other participants received the IMC after having completed the non-differentiation measure (constituting our IMC order manipulation). This was conducted to determine whether the IMC had an effect on the mindsets of the participants. Additionally, the order of the non-differentiation measure was varied, such that the non-differentiation task either appeared as the third question in the series of tasks or the sixth question in the series of tasks (the task order manipulation). This was done to see the effect of order on the tasks.

The IMC (see Appendix A; adopted from Oppenheimer et al., 2009) appeared on its own page in the survey, with "mark all" checkboxes for the different sports options. The title, "SPORTS PARTICIPATION" and the lure question, "Which of these activities do you engage in regularly?" were bolded and in large font, while the instructional paragraph was unbolded and in a standard font size. The response option marked "other" was followed by a textbox to allow for text entry.

The instructional paragraph for this question asked participants to disregard the standard response format for the section in lieu of a response that confirmed that they had read the instructions. Therefore, instead of responding to the lure question "Which of these activities do you engage in regularly?" by selecting one or more of the checkboxes detailing sports options, participants were asked to enter "I read the instructions" in the response option marked "other" before continuing with the survey. Thus, participants who entered "I read the instructions" in the box marked "other" were the ones considered to "pass" the IMC.

The non-differentiation measure followed a standard matrix rating scale format (see Appendix B). The question appeared on a separate page in the survey, and asked participants to rate their interest in reality TV shows (ten real, three fictitious) on the following scale: extremely interesting (1), very interesting (2), fairly interesting (3), not too interesting (4) and not interesting at all (5). Participants were also given the option to leave items blank, if they wished to express no opinion. Following standard procedure (Krosnick & Alwin, 1988) for each participant, the number of TV shows assigned the same rating was counted, and became the basis for their non-differentiation score. A score of thirteen indicated that the participant rated all thirteen TV shows the same, and thus displayed very high non-differentiation; whereas a score of three indicated that the participant only rated three qualities the same, and thus displayed low non-differentiation.

Results

Seven hundred and forty six participants (93%) responded correctly to the IMC. One participant did not complete the entire non-differentiation measure, leaving us with 745 participants for this analysis.

Non-differentiation scores were computed by counting the number of TV shows to which each participant assigned the same rating (Krosnick & Alwin, 1988). Thus, the more shows that were given the same rating, the more non-differentiation a participant displayed. Five hundred and eighty participants (77.7%) rated more items as "not interesting at all" (5) than any other option. Ninety-seven participants (13%) gave all 13 shows the same rating, while only 2 participants (< 1%) gave 3 shows the same rating. The mean non-differentiation score was 8.97, with a standard deviation of 2.74.

In order to assess the priming effect of the IMC on the non-differentiation measure, we conducted a 2 (IMC order: first, last) x 2 (task order: third, sixth) analysis of variance on non-differentiation scores. The sample was limited to only the 745 participants who responded correctly to the IMC, because the sample of participants who responded incorrectly (53 participants) was not large enough for drawing firm conclusions. Furthermore, participants who responded correctly are considered non-satisficers, and as such, are the most theoretically relevant population for the study, as the sole reason why researchers include an IMC in a survey is to whittle their sample down to just those participants (Oppenheimer et al., 2009).

There was a main effect of task order: $F(1, 741) = 8.15$, $p = .004$, $\eta_p^2 = .011$. The mean non-differentiation score was higher when the Reality TV Show task was presented sixth in the survey ($M = 9.26$, $SD = 2.66$), as compared to when it was presented third in the survey ($M = 8.69$, $SD = 2.79$). This suggests sensitivity to satisficing. Intuitively, however, this makes sense as fatigue is likely to set in as the survey progresses, and fatigued participants are more likely to satisfice (Krosnick, 1991).

Despite this sensitivity, we were unsuccessful in obtaining an IMC order effect. We had hypothesized that the presence of an IMC preceding the non-differentiation measure would lower

participant's non-differentiation score (indicating lower levels of satisficing), regardless of the order in which the non-differentiation measure appeared in the survey. However, there was no main effect of IMC order: $F > 1$. Thus, order of the IMC, either preceding or following the non-differentiation measure, had no statistically significant effect on participants' satisficing behaviors.

The interaction of IMC order and task order did not reach significance: $F(1, 741) = 1.91$, $p = .17$, $\eta_p^2 = .003$.

Small Discussion

Contrary to the hypothesis, the results of this study show that the inclusion of the IMC in this survey did not have an effect on participants' subsequent non-differentiation behavior. Instead, the order in which the task appeared in the survey packet was the only factor to have a significant impact on participants' responses. Intuitively, the effect of a task order makes sense, as lengthy surveys generally cause fatigue in participants that make them more likely to adopt satisficing behaviors as the survey progresses (Krosnick, 1991).

Study 2

Survey research has demonstrated that participants often infer the meanings of the questions and response choices presented in surveys by relying on the norms and expectations associated with everyday conversation (Krosnick, 1999; Schwarz, 1996). In conversation, speakers depend on a set of conventions regarding speech patterns and topics. This set of tacit assumptions can be expressed as Gricean norms (Grice, 1975; Schwarz, 1996), which not only increase efficiency, but allow speakers to convey unstated and unexpressed messages with their responses (Krosnick, 1999). In survey research, participants similarly rely on these Gricean norms when they encounter survey questions and formulate their responses (Krosnick, 1999;

Schwarz, 1996). Thus, participants often look beyond the literal meaning of the question presented, and instead, rely on the pragmatic meaning of the information to produce a response (Schwarz, 1996).

However, participants not only apply these Gricean norms to verbal content presented in survey research, but also to the formal features of the questionnaire (Schwarz, 1994). Schwarz (1994), for example, states that the overall set of the numerical values that are used in survey research (e.g. a scale from -5/+5 and a scale from 0/10) can have an influence on participants' responses. This is likely due to the differential interpretation of the numerical endpoints. Thus, a numeric value of 0 (on a 0 to 10 scale) suggests a unipolar dimension, whereby each scale value is a reflection of distinct degrees of the existence of the feature described in the question. On the contrary, a numeric value of -5 (on a -5 to +5 scale) suggests a bipolar dimension, implying that the negative label means the opposite of the described feature (Schwarz, 1994).

Other research has also recognized such effects, showing how participants draw inferences about how to answer questions from the numerical scale labels presented in a survey. Schwarz, Grayson, & Knäuper (1998), for example, showed how participants were likely to draw different inferences for the scale label "rarely" when it was combined with a numerical scale label of 0 versus when it was associated with a scale label of 1.

Schwarz (1994) further states that the range of response alternatives presented can serve as a source of information to participants, which subsequently influences their responses. This becomes especially problematic when researchers ask for reports of mundane behaviors, as these events are usually not represented as distinct episodes in memory (Schwarz, 1994). In such a scenario, participants have to rely on estimation, which may entail the use of the scale range to provide frequency estimates of their behavior. As participants assume that the scale they have

been presented with has been constructed using available information about the distribution of the specified behavior in the population (when it might not have), it can lead to inaccurate estimates of behavior frequency (Schwarz, 1994). Therefore, depending on the scale range participants are presented with (e.g. from 0.5 to 4.5 or more, or from 4.5 to 8.5 or more), the participant averages for the same question can differ significantly.

This shows how participants make systematic use of Gricean norms, in their attempt to understand the survey question and provide an appropriate response. Yet, as the conversations in research settings are usually unlike conversations in everyday settings (Schwarz, 1996), these studies have demonstrated the importance of shielding against the possibility of inadvertently imparting information to participants when violating conversational norms (Krosnick, 1999).

Existing literature in this field has shown that participants often infer meaning from minor aspects of questionnaire design (Schwarz, 1994). Presumably, this would be more likely to occur with increased attention to the wording of the question. Because an IMC essentially trains participants to pay close attention to minute aspects of a survey, we hypothesize that the presentation of the IMC preceding the Gricean norm effect measures will increase classic Gricean norm effects.

Method

Participants

Three hundred and ninety six participants on MTurk participated in the study (254 male, 142 female). The participants ranged in age from 18 to 50. The survey took between four to six minutes to complete, and each participant was compensated with 40 cents for participating.

Materials and Procedure

This study had a similar design to Study 1. Participants were first presented with a survey ostensibly on current issues. Participants completed a series of questions, including an IMC and an array of Gricean norm effect measures, but were randomly assigned to complete them in different orders. Half of the participants received the IMC as the first question in the survey, whereas, the other participants received the IMC after having completed the Gricean norm effect measures (the IMC order manipulation).

The IMC (see Appendix A; adopted from Oppenheimer et al., 2009) followed the same format as in Study 1. However, unlike Study 1, participants were also randomly assigned to receive feedback on their response. Feedback informed participants of incorrect answers and returned them back to the IMC with the instructions "Please try again" in the event of an incorrect response. Participants assigned to receive no feedback were not informed of incorrect answers, and thus simply progressed to the next page of the survey in the event of an incorrect response.

The Gricean norm effect measures included scale label and scale range effects on behavioral frequency judgments and judgments about one's life. The life success task (see Appendix C; adapted from Schwarz, Knäuper, Hippler, Noelle-Neumman, & Clark, 1991) asked participants, "How successful have you been in life so far?" on a 11 point scale anchored by "not successful at all" to "extremely successful." Participants were randomly assigned to one of two numerical scale labels. Half of the participants received the question with a scale ranging from 0, indicating "not successful at all," to 10, indicating "extremely successful," whereas the other half received it with a scale ranging from -5, indicating "not successful at all," to +5, indicating "extremely successful."

The rarely task (see Appendix D; adapted from Schwarz et al., 1998) asked participants, "How often do you get a haircut?," How often do you visit a museum?," and "How often do you attend a poetry reading?" with response options ranging from "rarely" to "often." However, as with the life success task, participants were randomly assigned to one of two numerical scale labels. Consequently, half of the participants received the question with a 0 to 10 rating scale, and half received it with a 1 to 11 rating scale.

The TV viewing task (see Appendix E; adapted from Schwarz, Hippler, Deutsch, & Strack, 1985) asked participants "On average, how many hours of TV do you watch daily?" As with the previous two tasks, participants were randomly assigned to one of two scale ranges. Therefore, half of the participants received the question with a scale ranging in frequency from up to 0.5 to 4.5 hours or more (low scale range), and half received it with the scale ranging in frequency from up to 4.5 to 8.5 hours or more (high scale range). The scale was created around the actual mean hours of TV viewed per day (4.5 hours; Nielsen, 2011), and thus both scale range conditions contained that mean (as the 2nd to last option in the low scale range, and the 2nd option in the high scale range). Following this question, participants were then asked, "How important is the role of TV in your leisure time?" with a scale option ranging from 0, indicating "not at all important," to 10, indicating "very important."

Results

We restricted our sample to only the 369 participants who responded correctly to the IMC, since the sample of participants who responded incorrectly was not large enough for drawing firm conclusions.

TV Viewing Task

Replication Analysis. The TV viewing task assesses whether the amount of people who say they watch more than the average amount of TV per day is moderated by the frequency range that is presented. The Gricean norm effect states that participants often infer that the middle point of the scale is the average of a measure in the general population. Today, the average amount of TV that people watch per day is 4.5 hours (Nielsen, 2011). Thus, to replicate the effects of past studies, more participants should have indicated watching more than the average amount of TV when presented with the high scale range. The analysis of this study showed that more participants estimated a higher than average TV consumption when presented with a high (rather than low) scale range. When participants were offered the high scale range, 19.6% of the participants reported watching more than the mean amount of TV per day, but when they were presented the low scale range, only 4.6% of participants reported watching more than the mean, χ^2 (1, $N = 369$) = 19.01, $p < .001$, $\Phi = .23$. Thus, we were able to replicate the effects of the original scale range task.

Does IMC order moderate scale range effects? We conducted a logistic regression with IMC order, feedback and scale range - and their interactions - entered as predictors of the likelihood of participants saying they watch more than the mean amount of TV per day. If IMC order influences Gricean norm effects, we would see an interaction between IMC order and the scale range presented. However, our results indicated that such an interaction was not significant, $B = -.50$, $Wald = .14$, $p = .71$. However, as the replication analysis shows, the effect of the scale range was significant in influencing the likelihood of reporting watching more than the mean number of hours of TV per day, $B = -1.70$, $Wald = 4.41$, $p = .04$, *odds ratio* = .18. All other main effects and interactions were not significant ($ps > .33$).

Typically, when participants get the low scale range, they are more likely to say that TV plays a more important role in their lives, as compared to when they get the high scale range. In order to investigate if IMC order moderates this effect of scale range on inferences of the importance of TV, we conducted a 2 (IMC order: first, last) x 2 (feedback: yes, no) x 2 (scale range: low, high) between subjects analysis of variance on the importance of TV in participants' lives. There were no main effects (ps > .10). However, we did get the hypothesized IMC order by scale range interaction, $F(1, 361) = 4.42$, $p = .04$. There was no significant difference in reports of how important TV was in the lives of participants receiving the low scale range ($M = 4.63$, $SD = 2.57$) and the high scale range ($M = 4.80$, $SD = 2.49$), when participants received the IMC first, $F(1, 361), = .43$, $p = .51$ for the simple effect. Yet, when they received the IMC last, there was the typical effect of scale range. Thus, those participants presented with a low scale range reported TV as being more important in their lives ($M = 5.38$, $SD = 2.41$), compared to those participants who received the high scale range ($M = 4.62$, $SD = 2.63$), $F(1, 361), = 5.48$, $p = .02$ for the simple effect.

Life Success Task

Replication Analysis. The life success task assess whether participants' rating of their success in life is moderated by different scale labels. The notion is that, when you present participants with the -5 to +5 scale range, they interpret the lower half values as describing failure, but when presented with the 0 to 10 scale range, the lower half is interpreted as the absence of success. Hence, participants are less likely to report ratings in the lower half of the scale range when presented with the -5 to +5 (Schwarz et al., 1991).

To replicate effects of the original study, fewer participants should have rated themselves with values in the lower half of the scale range when presented with the -5 to +5 (rather than the

0 to 10) scale. When participants were given the -5 to +5 scale labels, 35.2% reported a life

success rating between -5 and 0 (the lower half of the scale). Whereas, when participants were

given the 0 to 10 scale labels, 45.1% reported a life success rating in the lower half of the scale,

χ^2 (1, $N = 369$) = 3.71, $p = .05$, $\Phi = .10$. Thus, we were able to replicate the standard effects of

the original study.

Does IMC order moderate the scale label effect? We conducted a logistic regression,

where IMC order, feedback, scale labels - and their interaction - were entered as predictors of

participants' placing themselves in the lower half of the life success scale. If an IMC makes

participants focus more on minute aspects of the questions, it is expected that IMC order will

moderate the effect of scale label on the likelihood of a participant placing him/herself in the

lower half of the success scale. However, none of the main effects or interactions were

significant (ps > .26).

Rarely Task

The rarely task assessed whether participants frequency ratings of partaking in different

activities are influenced by different scale labels. We conducted a 2 (IMC order: first, last) x 2

(feedback: yes, no) x 2 (scale label: 0 to 10, 1 to 11) x 3 (item: haircut, museum, poetry reading)

mixed factorial analysis of variance. If IMCs influence participants to engage in deeper thought,

then it is expected that there will be some interaction between the IMC order and the scale labels

presented. However, neither the prior effects, nor the moderation by the IMC order, were

significant (all ps > .17).

<div align="center">**Small Discussion**</div>

Contrary to the hypothesis, the results of this study show that the inclusion of the IMC in

this survey did not moderate standard effects on either of the scale labeling tasks (i.e. the life

success task and the rarely task). The behavioral frequency task (i.e. the TV viewing task) provided curious findings; the IMC order again did not moderate the standard Gricean norm effect of scale range on behavior frequency estimates. However, the IMC order did moderate the effect of scale range on the question regarding the importance of TV in the life of participants. Thus, when participants received the IMC last, those who were presented with the low scale range reported TV as being more important, compared to those who received the high scale range.

General Discussion

Participants in online studies often fail to take the study seriously, and subsequently, forgo reading instructions and deeply processing the question (Oppenheimer et al., 2009). Hence, measures to identify satisficing participants become important for increasing the statistical power of the study. However, research on survey design has suggested that participants can infer intentions from prior questions, thereby affecting their responses to later questions (Strack et al., 1988). Hauser and Schwarz (2013a), for example, found that the inclusion of the IMC affected participants' responses on several variables associated with increased error processing. Hence, their research suggested that the IMC acts not just as a measure, but as a manipulation that changes the mindset of participants. This research laid the foundation for the notion that the inclusion of IMC's in survey research might also force the activation of deeper thought in participants, and thus, manipulate participants' survey responses.

Consequently, this study sought to identify whether the presentation of an IMC preceding both a satisficing measure and several Gricean norm measures would modify participants' behavior. If IMCs cause people to think more in a task, then it would decrease satisficing behaviors, as satisficing results from participants' desire to reduce cognitive demands. At the

same time, if IMCs cause participants to pay close attention to minute aspects of a survey, then it would increase classic Gricean norm effects.

However, this study shows that the inclusion of an IMC preceding a non-differentiation measure does not have an effect on participants' subsequent survey responses. Thus, it might just be that IMCs cause people to pay more attention to a task, but do not actually cause people to think more. Yet, the inclusion of the IMC did not moderate standard effects on both scale labeling tasks, nor the standard Gricean norm effect of scale range on behavior frequency either. Though, the IMC order did moderate the effect of scale range on the question regarding the importance of TV in the life of participants, but only when presented with the IMC last.

Because IMCs tell participants of a communicative context wherein the experimenter intends to trick participants into giving wrong answers, it is possible that, in subsequent questions, participants continue to expect violations in the rules of communication. If this is the case, it would explain why participants who received the IMC prior to the importance of TV question did not show standard Gricean norm effects. Nonetheless, if trust (or its lack thereof) is what explains this moderation of behavior, then there should have also been a moderation in participants' inferences on the frequency of their TV viewing by IMC order.

Consequently, this moderation (between the IMC and the importance of TV) might not be the result of trust, but a result of different processing styles. Local processing occurs when individuals focus on one feature of the task, while ignoring outside influences; whereas, global processing involves incorporating each element of the task into a whole to produce a complete picture (Navon, 1977). As IMCs convey to participants that there may be more to questions than meets the eye, they are likely to trigger local processing. If subsequent questions rely on local processing, then responses to those questions would not be further moderated by the IMC. The

question on frequency estimates of TV viewing behavior is the first question asked, and thus, requires only local processing to answer. This, therefore, would explain why there was no influence of the IMC on participants' responses to this question.

However, because the presentation of the IMC triggers local processing, it causes participants to consider questions in a vacuum, without relying on knowledge from preceding questions, and thereby impeding global processing. The question on the importance of TV involves global processing, as participants generally answer it by relying on information from the first question. However, when presented with the IMC first, this global processing is interrupted, and thus would explain why participants who received the IMC preceding this question showed no standard Gricean norm effects.

In either case, more research still has to be done on the influence of IMC's before researchers can unequivocally rely on them. Although this study shows that IMCs do not prompt additional cognitive processing, the curious moderation of TV importance and IMC order suggests that there may be some influence of IMCs that we have yet to understand. Hence, it may be necessary to conduct further research, using more well-known tasks, to assess whether IMCs have an influence on other cognitive processing skills, and whether this influence has any effects on survey results.

References

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic*

    Perspectives, 19(4), 25-42. doi: 10.1257/089533005775196732

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan, *Syntax and semantics*

    III: Speech acts (pp. 41-58). New York, NY: Academic Press.

Hauser, D. J., & Schwarz, N. (2013a). *Consequences of trap questions on survey respondent*

    mindsets. Manuscript in preparation.

Hauser, D. J., & Schwarz, N. (2013b). *Effects of IMCs on error detection processing*.

    Unpublished manuscript, Department of Psychology, University of Michigan, Ann

    Arbor, USA.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude

    measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. doi:

    10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567. doi:

    10.1146/annurev.psych.50.1.537

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis:

    Ratings, rankings and the measurement of values. *Public Opinion Quarterly, 52*(4), 526-

    538. doi: 10.1086/269128

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception.

    Cognitive Psychology, 7, 353-383. doi: 10.1016/0010-0285(77)90012-3

Nielsen. (2011). *State of the media: The cross-platform report.* New York, NY: The Nielsen

    Company.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks:

Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872. doi:10.1016/j.jesp.2009.03.009

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna, *Advances in experimental social psychology* (Vol. 26, pp. 123-162). San Diego, CA: Academic Press.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation.* Hillsdale, NJ: Erlbaum.

Schwarz, N., Grayson, C. E., & Knäuper, B. (1998). Formal features of rating scales and the interpretation of question meaning. *International Journal of Public Opinion Research, 10*(2), 177-183. doi: 10.1093/ijpor/10.2.177

Schwarz, N., Hippler, H., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *The Public Opinion Quarterly, 49*(3), 388-395. doi: 10.1086/268936

Schwarz, N., Knäuper, B., Hippler, H., Noelle-Neumman, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *The Public Opinion Quarterly, 55*(4), 570-582. doi: 10.1086/269282

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition, 26*(6), 791–799. doi: 10.1521/soco.2008.26.6.791

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37*(10), 1660-1672. doi: 10.1037/0022-3514.37.10.1660

Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social

determinants of information use in judgments of life satisfaction. *European Journal of*

Social Psychology, 18, 429-442. doi: 10.1002/ejsp.2420180505

Appendix A

SPORTS PARTICIPATION

Most modern theories of decision making recognize the fact that decisions do not take place in a

vacuum. Individual preferences and knowledge, along with situational variables can greatly

impact the decision process. In order to facilitate our research on decision making we are

interested in knowing certain factors about you, the decision maker. Specifically, we are

interested in whether you actually take the time to read the directions; if not, then some of our

manipulations that rely on changes in the instructions will be ineffective. So, in order to

demonstrate that you have read the instructions, please ignore the sports items below. Instead,

select the box marked "other" and type "I read the instructions" (no quotes) in the textbox, then

click continue. Thank you very much.

"Which of these activities do you engage in regularly? (click on all that apply)"

☐ skiing

☐ soccer

☐ snowboarding

☐ running

☐ hockey

☐ football

☐ swimming

☐ tennis

☐ basketball

☐ cycling

☐ other: (text entry)

Appendix B

REALITY TV SHOWS

Please look at the reality television shows listed below. Could you please tell me whether you find the reality television show to be extremely interesting, very interesting, fairly interesting, not too interesting, or not interesting at all.

1.   The Real Teenagers of Beverly Hills

2.   Survivor

3.   Fish Tank Kings

4.   The Biggest Loser

5.   Hell's Kitchen

6.   So You Think You Can Dance?

7.   Shahs of Sunset

8.   Geeks vs. Greeks

9.   Married to a Vampire

10.   America's Next Top Model

11.   Millionaire Matchmaker

12.   The Bachelor

13.   The Apprentice

Scale: *Extremely interesting (1)*, *Very interesting (2)*, *Fairly interesting (3)*, *Not too interesting (4)* and *Not interesting at all (5)*

Appendix C

LIFE SUCCESS

How successful have you been in life so far? Please use the following rating scale (0 means that you have been 'not successful at all,' and 10 means that you have been 'extremely successful').

Scale: *Not successful at all (0)* to *Extremely successful (10)*

[OR]

How successful have you been in life so far? Please use the following rating scale (-5 means that you have been 'not successful at all,' and +5 means that you have been 'extremely successful').

Scale: *Not successful at all (-5)* to *Extremely successful (+5)*

Appendix D

RARELY/OFTEN

How often do you get a haircut?

Rarely 0 1 2 3 4 5 6 7 8 9 10 Often

How often do you visit a museum?

Rarely 0 1 2 3 4 5 6 7 8 9 10 Often

How often do you attend a poetry reading?

Rarely 0 1 2 3 4 5 6 7 8 9 10 Often

[OR]

How often do you get a haircut?

Rarely   1 2 3 4 5 6 7 8 9 10 11 Often

How often do you visit a museum?

Rarely 1 2 3 4 5 6 7 8 9 10 11 Often

How often do you attend a poetry reading?

Rarely 1 2 3 4 5 6 7 8 9 10 11 Often

Appendix E

DAILY TV CONSUMPTION

On average, how many hours of TV do you watch daily?

O   Up to ½ h

O   ½ h to 1 ½ h

O   1 ½ h to 2 ½ h

O   2 ½ h to 3 ½ h

O   3 ½ h to 4 ½ h

O   More than 4 ½ h

[OR]

On average, how many hours of TV do you watch daily?

O   Up to 4 ½ h

O   4 ½ h to 5 ½ h

O   5 ½ h to 6 ½ h

O   6 ½ h to 7 ½ h

O   7 ½ h to 8 ½ h

O   More than 8 ½ h

[AND]

How important is the role of TV in your leisure time?

Scale: *Not at all important (0)* to *Very important (10)*