Journal of Experimental Social Psychology 45 (2009) 867–872

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Reports

Instructional manipulation checks: Detecting satisficing to increase statistical power

Daniel M. Oppenheimer [a,*], Tom Meyvis [b], Nicolas Davidenko [c]

[a] Princeton University, Dept. of Psychology, Green Hall, Princeton, NJ 08540, USA
[b] Stern School of Business, New York University, 8–15 Tisch Hall, New York, NY 10012, USA
[c] Psychology Department, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 19 August 2008
Revised 26 March 2009
Available online 5 April 2009

Keywords:
Satisficing
Instructions
Screening
Manipulation checks

ABSTRACT

Participants are not always as diligent in reading and following instructions as experimenters would like them to be. When participants fail to follow instructions, this increases noise and decreases the validity of their data. This paper presents and validates a new tool for detecting participants who are not following instructions – the Instructional manipulation check (IMC). We demonstrate how the inclusion of an IMC can increase statistical power and reliability of a dataset.

© 2009 Elsevier Inc. All rights reserved.

Most experimenters have dealt with participants who are not as diligent as we would like them to be. Some participants, who give flippant answers or whose answers fall many standard deviations from the mean are discarded as outliers on the assumption that the data are purely noise and will merely serve to dilute the signal. However, not all participants advertise their negligence so blatantly. Some participants may skim instructions, missing key elements of the task or manipulation, or respond in a haphazard fashion that defies outlier analysis. These participants increase noise, reduce experimental power, and thus force experimenters to expend resources running more participants than would otherwise be necessary. Unfortunately, such participants have traditionally been challenging to detect and account for.

One theory that has addressed this problem is Krosnick's (1991) theory of satisficing in survey responses. Krosnick built on Herb Simon's (1957) idea that people have limited cognitive resources and attempt to minimize cognitive effort. To reduce demand, participants might satisfice: rather than attempting to find an optimal solution to a problem, people might go with the first minimally acceptable alternative that comes to mind.

Krosnick noted that responding to surveys often requires a great deal of cognitive effort (Krosnick, 1991; Krosnick, Narayan, & Smith, 1996). He hypothesized that participants might satisfice by choosing the first (as opposed to *best*) alternative that fits the question or, in extreme cases, by answering randomly.

By providing answers that do not accurately address the questions, satisficing participants decrease the signal-to-noise ratio of a data set, and can substantially lower the power of an experiment. We propose that participants who are satisficing will often not bother to read the questions or instructions in a survey. Assuming that these questions and instructions are necessary to enable participants to complete the survey in a way that produces useful data, identifying these participants could substantially increase the power of the study.

Therefore, to detect satisficers we have developed a new methodological tool: the Instructional manipulation check (IMC). The IMC measures whether or not participants are reading the instructions, and thus provides an indirect measure of satisficing. It consists of a question embedded within the experimental materials that is similar to the other questions in length and response format (e.g. Likert scale, check boxes, etc.). However, unlike the other questions, the IMC asks participants to ignore the standard response format and instead provide a confirmation that they have read the instruction. For an example, see Fig. 1.

We gave a paper and pencil IMC in a packet of unrelated questionnaires to three different samples of participants. The text was identical to that in Fig. 1, with the exception that participants were told to "write 'I read the instructions' somewhere on the page" rather than clicking the title. To test an inherently motivated sample, we recruited 336 Stanford University undergraduates who were considering either a major or a minor in psychology and thus were expected to be motivated to take the survey seriously. To test a less motivated sample, we recruited 87 Stanford University undergraduates and visiting high school students who were not considering a major or minor in psychology. These participants were expected to be relatively unmotivated and more likely to engage in satisficing. Finally, to see if we could improve motivation, we recruited a third sample of 57 Stanford University

* Corresponding author.
E-mail address: doppenhe@princeton.edu (D.M. Oppenheimer).

0022-1031/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2009.03.009

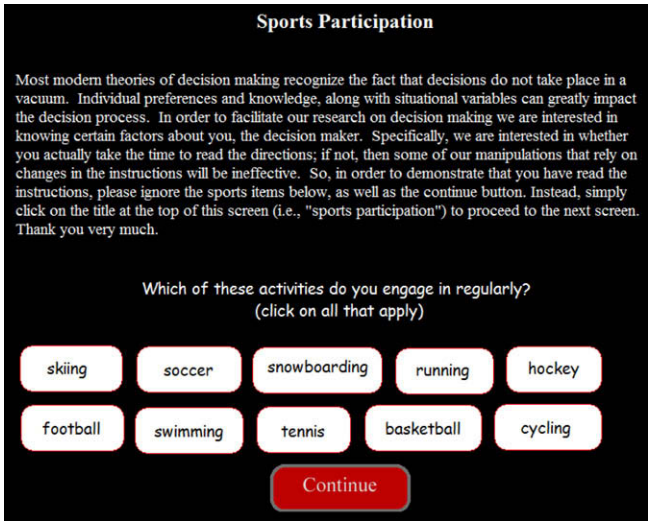

Fig. 1. Screen shot of IMC.

undergraduates who, like the second sample, were not considering a major or a minor in psychology and thus were expected to be relatively unmotivated. However, these surveys were proctored; supervision was expected to lead participants to take the surveys more seriously.

Participants who responded correctly by writing, "I read the instructions," or some variant thereof, were coded as having passed the IMC. Participants who failed to write anything about the instructions, or who filled out the activities preference form were coded as having failed the IMC.

As expected, the highest failure rate (28.7%) occurred in the non-motivated sample. This failure rate was a significantly higher than in both the motivated condition (17.5%; $\chi^2(1) = 5.42$, $p < .05$) and the supervised condition (14.0%; $\chi^2(1) = 4.21$, $p < .05$).

These preliminary findings demonstrate that a substantial number of people do not follow instructions when filling out questionnaires. Consistent with our interpretation of the IMC as a measure of satisficing, motivated or supervised participants were less likely to fail the IMC than participants who were neither motivated nor supervised.

Data from non-diligent participants adds noise and can substantially decrease statistical power. By including an IMC in a study, one can potentially identify this source of noise and eliminate it, thereby increasing power and allowing for reliable results with fewer resources and participants. Estimating the increase in power allowed through use of an IMC rests on the following assumptions: (1) participants who fail the IMC also fail to follow other instructions in the survey; and (2) failing to follow these other instructions will result in less reliable and valid data. Study 1 tests these assumptions.

Study 1

Study 1 investigated whether using an IMC to detect satisficing participants actually reduces noise and increases the power of an experiment. To this end, we replicated two well established and robust paradigms from the judgment and decision making literature, and examined whether participants who failed the IMC also failed to show the standard effects. We also examined whether there were systematic differences between participants who passed and failed the IMC, including demographics, self-reported motivation, or Need for Cognition (Cacioppo, Petty, & Kao, 1984).

Method

Participants

Two hundred and thirteen participants (156 women, 57 men) from New York University took part in the study in partial fulfillment of a course requirement or in exchange for $10.

Stimuli, procedure and design

To determine whether removing participants who failed the IMC would increase the power of studies, two classic paradigms from the judgment and decision making literature were replicated. The first was Thaler's (1985) beer pricing task that demonstrates how different expectations can change people's willingness to pay for identical experiences. As many of the participants were under the legal drinking age, the scenario was changed to soda pricing. The exact wording was as follows:

> You are on the beach on a hot day. For the last hour you have been thinking about how much you would enjoy an ice cold can of soda. Your companion needs to make a phone call and offers to bring back a soda from the only nearby place where drinks are sold, which happens to be a [run-down grocery store] [fancy resort]. Your companion asks how much you are willing to pay for the soda and will only buy it if it is below the price you state. How much are you willing to pay?

Participants randomly received one of the two versions of the scenario. Thaler (1985) found that participants were willing to pay substantially more for a beer from a fancy resort than from a run down grocery store, even though the experience of drinking the beer would be identical regardless of the source. However, the differences between the scenarios are quite subtle, consisting of changes to only two or three words. Therefore, a participant who was not paying close attention to the question may not be effectively exposed to the manipulation, and would thus serve as a source of noise.

The second paradigm was a sunk cost question, also adapted from Thaler (1985). The exact wording was as follows:

> Imagine that your favorite football team is playing an important game. You have a ticket to the game that you [have paid handsomely for] [have received for free from a friend]. However, on the day of the game, it happens to be freezing cold. What do you do?

Participants randomly received one of the two versions of the scenario and indicated their intention to attend the game on a nine point scale (1 = *definitely stay at home*, 9 = *definitely go to the game*). Previous research has found that people are less likely to skip the game if they have paid for the tickets (Arkes & Blumer, 1985; Thaler, 1980). However, the effect relies on subtle differences in the wording that may be overlooked by participants who are not reading the questions carefully. Following the two judgment tasks, participants were provided with an IMC as described in the introduction (see Fig. 1 for a screen shot of the IMC).

We next examined how participants who failed the IMC differed from those who passed it. Participants provided demographic information and filled out an abbreviated 18 item Need for Cognition scale (Cacioppo et al., 1984). Furthermore, we recorded the amount of time that each participant took to complete the study, as participants who satisfice should take less time. Finally, participants indicated how motivated they were to complete the study (1 – *not motivated at all*, 9 – *very motivated*).

Results

Participants who clicked on the sports categories or the continue button rather than the title (as they were instructed to do),

Table 1
Examples of domains in which manipulations often consist of small differences in instructions.

Domain	Example study
Game theory	Liberman, Samuels, and Ross (2004): Changing the name of a prisoner's dilemma scenario ("*Wall St. Game*" vs. "*Community Game*") affected willingness to cooperate or defect
Clinical psychology	Page and Bennesch (1993): People rated themselves as more depressed on clinical tests when the instructions indicated they were participating in a "*Hassles of living*" questionnaire vs. a "*Depression*" questionnaire
Memory	Loftus and Palmer (1974): After watching a video of a car crash, people recalled the car as moving faster when the question was asked using the verb "*smashed*" in place of "*collided*" or "*bumped*"
Impression formation	Kelley (1950): Students rated a guest instructor more favorably when, prior to class, they read a short bio which described him as "*very warm*" as opposed to "*rather cold*"
Regulatory fit	Lee and Aaker (2003): Participants rated a brand higher when a gain frame was paired with a promotion tagline (such as "*Get Energized!*"), or a loss frame was tagged with a prevention tagline (such as "*Do not miss out on getting energized!*")
Prejudice/stereotyping	Bertrand and Mullainathan (2004): Changing the name on a resume from "*Emily*" or "*Greg*" to "*Lakisha*" or "*Jamal*" affected the applicant's chance of being hired
Judgment and decision making	Thaler (1985): People are willing to pay more for a beer when it was described as being purchased from a "*fancy resort hotel*", rather than a "*run-down grocery store*"
Language	Glucksberg and Haught (2006): found that similes and metaphors (such as "*Obsession is a wart*" vs. "*Obsession is like a wart*") can dramatically differ in the way we mentally interpret them (categorization vs. comparison)
Attitudes	Salancik and Conway (1975): People's attitudes towards religion were influenced by filling out a questionnaire with statements such as "*I occasionally cheat on examinations*" vs. "*I frequently cheat on examinations*"
Moral reasoning	Monin and Miller (2001): People gain more moral credentials by filling out a true–false questionnaire with statements such as "*Most women are not really smart*" than "*Some women are not really smart*"

were coded as having failed the IMC. Overall, 46% of the sample failed the IMC.

Next, we examined if taking participants' performance on the IMC into account would improve our ability to reveal the price expectation and sunk cost effects observed by Thaler (1985). The initial analysis of the soda pricing data indicated that we failed to replicate the effect of the source manipulation, $F(1, 211) = 0.25$, $p > .1$. However, after restricting the data set to those participants who passed the IMC, we observed that people were in fact willing to pay significantly more when the source was a resort ($M = \$3.04$) than when the source was a grocery store ($M = \$2.05$), $F(1, 113) = 7.72$, $p < .01$, $\eta_p^2 = .06$. Participants who failed the IMC showed no significant difference in what they would pay at the resort ($M = \$2.43$) and at the store ($M = \$3.05$), $F(1, 96) = 0.34$, $p > .1$. While there was no interaction between passing the IMC and price expectation ($F(1, 208) = 2.37$, $p = .125$), this was likely due to the fact that the variance for those who failed the IMC was quite large ($MSE_{PASS} = 3.52$, $MSE_{FAIL} = 26.37$), which supports the claim that participants who fail the IMC greatly increase noise.

In the sunk cost task, participants were slightly more likely to attend the game if they had paid for the ticket ($M = 7.46$) than if the ticket had been free ($M = 6.93$), $F(1, 211) = 2.74$, $p = .1$, $\eta_p^2 = .01$. However, this effect was more pronounced when the data set was restricted to those who passed the IMC, in spite of the considerable decrease in sample size ($M_{PAID} = 7.75$, $M_{FREE} = 6.93$), $F(1, 113) = 3.55$, $p = .06$, $\eta_p^2 = .03$. This occurred since, as in the soda study, those who failed the IMC did not show any effect of the manipulation ($M_{PAID} = 7.15$, $M_{FREE} = 6.93$), $F(1, 96) = 0.21$, $p > .1$. Again, the interaction between passing the IMC and the manipulation of interest was not reliable ($F(1, 209) = 0.88$, $p = .349$).

Study 1 also tested for individual differences between those who failed the IMC and those who passed it. There were no reliable differences in the IMC failure rate by age, gender, or whether the participants were participating in exchange for money or course credit. However, consistent with our interpretation of the IMC as a measure of satisficing, people who failed the IMC spent significantly less time on the study ($M = 153$ s) than those who passed the IMC ($M = 191$ s), $F(1, 211) = 42.86$, $p < .001$, $\eta_p^2 = .17$.[1]

Furthermore, participants who failed the IMC had significantly lower Need for Cognition scores ($M = 14.66$) than those who passed ($M = 25.15$), $F(1, 211) = 12.59$, $p < .001$, $\eta_p^2 = .06$. Interestingly, participants who failed the IMC also responded more inconsistently on the Need for Cognition scale than those who passed the IMC. As this scale is known to be internally consistent, participants' answers on the positive items should be negatively correlated with the answers on the reverse-scored items (Cacioppo et al., 1984). For participants who passed the IMC, the correlation was $r = -.57$, while for participants who failed the IMC, the correlation was reliably attenuated ($r = -.28$; $z = 2.64$, $p < .01$). Thus, the difference we found in Need for Cognition scores may be due, at least in part, to greater measurement error in the participants who failed the IMC.

Finally, we inspected participants' self-reported motivation. Participants who failed the IMC reported statistically the same level of motivation ($M = 5.5$) as those who passed ($M = 5.6$), $F(1, 211) = 0.31$, $p > .1$. The lack of difference in self-reported motivation could have many reasons, including self-presentation concerns, random responses, or a lack of insight into one's own motivation. Regardless of the explanation, the absence of a difference in self-reported motivation suggests that self-reports may not be a reliable way to detect satisficing in experiments.

Discussion

Study 1 demonstrated that the use of an IMC can substantially increase the statistical power of an experiment. In the soda pricing task, the overall results would lead one to conclude that there was no effect of price expectations on people's willingness to pay. However, when considering only the participants who passed the IMC, the effect was reliable and compelling. Similar results were found for the sunk cost effect. Despite the reduction in sample size, the power of the experiments was increased by eliminating participants who failed the IMC. While the interaction between passing the IMC and the variables of interest was not reliable, this may have been due to the tremendous variance observed among participants who failed the IMC. It is worth noting that the IMC could not influence responses to these target questions because the target questions temporally preceded the IMC.

Participants who failed the IMC took less time to complete the experiment and were reliably lower in Need for Cognition than those who passed. However, it is difficult to draw strong conclusions from this, as additional analysis indicated that the data collected from

[1] Time was recorded from experiment onset to experiment completion. In other words, measurements include the time spent reading instructions in addition to time spent on the key experimental tasks. However, in subsequent studies participants took an average of 24.7 s ($\sigma < 10$ s) to complete an IMC. Even if this were subtracted from the 38 s difference observed in the Study 1, the results would be reliable.

participants who fail the IMC was also less reliable: Participants who failed the IMC were less likely to distinguish between positive and negative (reverse coded) items, likely because they were not reading the items or scales carefully. As such, it is unclear to what extent the differences in Need for Cognition scores were due to real personality differences rather than measurement error. It is worth noting that the challenges in measuring Need for Cognition among participants who failed the IMC provide a third demonstration (in addition to the soda pricing and sunk cost paradigms) that such participants increase noise and reduce power in experiments.

Finally, participants who failed the IMC did not report being less motivated than those who passed. However, as with the Need for Cognition scale, it is difficult to draw conclusions from this result, as the self-report data from participants who fail the IMC is suspect. In fact, participants who failed the IMC did take reliably less time to complete the experiment than those who passed, indicating that there may indeed be motivational differences. As such, our results suggest that an IMC may be more effective than self-report measures at detecting satisficing and identifying problem subjects.

So far, we have suggested that participants who fail the IMC are a source of noise and that eliminating them will increase the power of experiments. However, one question that naturally arises when eliminating a subset of participants from a sample is whether this elimination will bias results. Krosnick et al. (1996) demonstrated that satisficing is associated with individual differences in cognitive load, and correlates to some extent with personality constructs as well. One could plausibly argue that removing these participants from the pool introduces generalizability issues. Perhaps participants who fail the IMC failed to react to the context manipulations in Study 1 not because they failed to notice these changes, but because they do not take price expectations or sunk costs into account. If so, eliminating them would be eliminating a true source of variance.

To test this possibility, we actively intervened on participants who failed the IMC to force them to read the instructions. Specifically, upon failing an IMC, participants were not allowed to continue in the experiment, but rather were prompted to "try again" until they read the instructions closely enough to pass. We hypothesize that the population that initially fails the IMC, once prompted to actually read the instructions, will subsequently behave like the population that passes the IMC.

Study 2

Method

Participants

One hundred and forty-four participants (76 women, 68 men) from New York University took part in the study in partial fulfillment of a course requirement or in exchange for $10.

Stimuli, procedure and design

Study 2 adopted the same procedure as Study 1, except that the IMC was presented at the beginning of the study, and participants who failed the IMC were not allowed to continue in the experiment, but were redirected to the IMC and forced to try again until they passed. Participants completed the same measures as in Study 1, with the addition of a 7-item abbreviated version of the Schwartz et al. (2002) maximization scale, which measures participants' tendency to maximize versus satisfice.

Results

Overall, 50 of the 144 participants (35%) failed the IMC. Of those who failed, 35 correctly responded on the second try, 9 on the third try, 5 on the fourth try, and 1 on the fifth try. There were no reliable demographic differences (age, gender, etc.) correlated with failure rate.

Unlike in Study 1, we now observed a reliable effect of the source manipulation across all participants. People were willing to pay more for a soda from a resort (M = $3.04) than from a grocery store (M = $2.06), $F(1, 134) = 25.65$, $p < .0001$, $\eta_p^2 = .16$. This was not only the case when we restricted the data to participants who initially passed the IMC (M_{RESORT} = $2.96, M_{STORE} = $2.02, $F(1, 88) = 15.87$, $p < .0001$, $\eta_p^2 = .15$), but also for participants who initially failed the IMC (M_{RESORT} = $3.23, M_{STORE} = $2.13, $F(1, 46) = 11.48$, $p < .005$, $\eta_p^2 = .20$).

A similar pattern of results was found for the sunk cost manipulation. The overall reported likelihood of attending the game was significantly higher when participants had paid for the ticket ($M = 8.04$) than when they had not ($M = 7.27$), $F(1, 140) = 5.88$, $p < .05$, $\eta_p^2 = .04$). The same trends were observed when we constricted the data set to those who passed the IMC ($M_{PAID} = 7.92$, $M_{FREE} = 7.29$, $F(1, 92) = 2.62$, $p = .1$, $\eta_p^2 = .03$), or to those who failed the IMC ($M_{PAID} = 8.23$, $M_{FREE} = 7.21$, $F(1, 48) = 3.26$, $p = .08$, $\eta_p^2 = .06$).

In contrast to Study 1, participants who passed the IMC did not spend more time on the task (M = 238 s) or show higher Need For Cognition scores (mean NFC = 16.67) than those who initially failed the IMC (M = 245 s, mean NFC = 18.61). Further, unlike in Study 1, correlations between positive and negative (reverse scored) items on the Need for Cognition Scale were not stronger for those who passed the IMC ($r = -.62$) than for those who initially failed the IMC ($r = -.69$), but participants who passed the IMC did report higher motivation ($M = 5.31$) than those who initially failed the IMC ($M = 4.22$), $F(1, 141) = 8.43$, $p < .01$, $\eta_p^2 = .06$. In addition, participants who passed the IMC scored marginally higher on maximizing (and thus lower on satisficing, $M = 6.76$) than those who failed the IMC ($M = 6.38$), $F(1, 136) = 3.28$, $p = .07$, $\eta_p^2 = .02$.

Discussion

After being forced to read the instructions, participants who initially failed the IMC became indistinguishable from those who initially passed on both the soda-pricing and sunk cost tasks. This strongly suggests that the differences observed in Study 1 were due to noisy responses by non-diligent participants, rather than by a sub-population that was inherently less sensitive to context effects. In addition, Need for Cognition scores were not reliably different between those who initially passed and those who initially failed the IMC, further reinforcing the notion that the differences observed in Study 1 were due to noisy responses. In other words, at least with regard to context effects, using an IMC does not remove a source of true variance in the population – participants who initially fail the IMC behave just like those who initially pass, provided they actually read the instructions. It is worth a cautionary note, however, that our measurement of personality differences was hardly exhaustive – other personality variables such as conscientiousness might differentiate amongst participants who pass and those who fail the IMC, and would be worth future study.

It is also worth noting that a difference emerged in self-reported motivation and satisficing/maximizing. Participants who initially failed the IMC also reported being less motivated and more likely to satisfice. Importantly, forcing participants to read the instructions did not impair measurement of the differences between groups.

General discussion

Not all participants take psychological research as seriously as we would like. In this paper, we demonstrated that a substantial

proportion of participants fail to follow instructions when completing surveys. Further, we have introduced a new tool – the Instructional manipulation check – which allows researchers to identify these participants. Eliminating participants who are answering randomly – or better yet, forcing these participants to read instructions more carefully – will increase the signal-to-noise ratio, and in turn increase statistical power.

The effectiveness of an IMC at increasing power depends on several factors. First, it depends on the proportion of a particular population engaging in satisficing. In these studies, IMC failure rates ranged from 14% to 46%. An IMC is more beneficial when the population contains more satisficers or when the nature of the task discourages diligent reading (e.g. instructions resembling the 'fine print' from legalistic documents and User License Agreements, etc.).

Second, the effectiveness of an IMC assumes that people who fail the IMC will also fail to read other instructions. In the studies above, the IMC was similar in form and style to other survey questions. As the differences between the IMC and the other questions increases, so too will the discrepancy between reading the IMC and reading the other instructions. To make sure this assumption is satisfied, the format of the IMC should be adjusted to match the format of questions in the rest of the survey.

One example of such an adjustment is the blue dot task shown in Fig. 2. The blue dot IMC takes the form of a Likert Scale. The IMC reads, "Please click on the little blue circle at the bottom of the screen. Do not click on the scale items that are labeled from 1 to 9," and is followed by a Likert Scale with endpoints of "*very rarely*" to "*very frequently*". In a sample of over 1000 undergraduate participants, approximately 7% fail this task, and click on the Likert Scale despite the fact that the scale makes no sense in the absence of a question. This serves as a demonstration of how IMCs can be varied to fit a particular study, and also how even for short, simple prompts, there is a non-negligible proportion of the population that is apparently answering at random.

Third, the effectiveness of an IMC is moderated by the extent to which satisficing leads to impaired performance. Some surveys do not require participants to read the directions in order to successfully answer the questions. For example, demographic information surveys can typically be answered without reading precise instructions. However, the effectiveness of participants at answering surveys without instructions relies critically on the extent to which their default interpretations of the task align with the goals of the experimenter. For example, if a researcher were to ask participants to report demographic information about their best friend, a person who had not read the instructions might provide answers corresponding to the wrong person. In cases where the effectiveness of a manipulation relies on subtleties within the instructions, such as the pricing and sunk cost paradigms used in this study, or in cases where the task is not self evident, an IMC would be particularly important. It is worth noting that an impressive array of studies across a wide range of psychological sub-disciplines make use of subtle (one or two word) variations in instructions as primary manipulations (see Table 1).

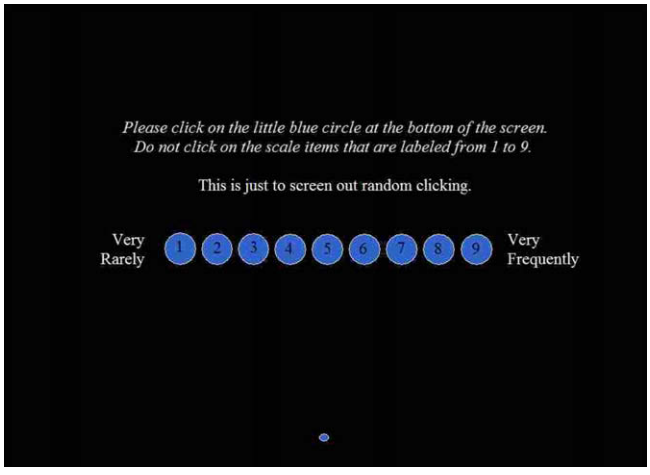

Fig. 2. Example of an alternate IMC – the Blue Dot task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Caveats

As with any instrument, the IMC has its drawbacks. Foremost, there is the concern that if an IMC is used to eliminate participants from the sample then the external validity of the study could be harmed. If the population that failed the IMC differed substantively from those who passed the IMC it could lead to issues regarding generalizability of the findings. While the present studies found no differences in passing rates based on age, gender, or need for cognition, that is hardly an exhaustive list of group differences. It may be that testing IMCs on a more diverse sample of participants may yield important differences between those who pass and those who fail the IMC (e.g. culture, education level, etc.). Further, while we found that participants who failed the IMC showed the same biases as diligent participants (upon being forced to read the instructions), we cannot guarantee that this would be the case across all domains listed in Table 1.

Another concern with using the IMC is the possibility of participant backlash. Diligent participants who come across an IMC may feel insulted to find that they are not trusted by the researchers. Satisficing participants may be embarrassed at having failed the IMC and may seek retribution for their embarrassment by trying to foil the study. There is also the concern that by including an IMC, researchers are, through pragmatic implicature, suggesting a norm of non-diligence. In other words, there is the possibility that including an IMC will hurt the quality of the data. However, in the reported studies, we found no evidence for such a backlash. Through the hundreds of participants that have taken part in IMC studies, there has yet to be a single complaint. Moreover, participants in Study 2 showed the classic effects after exposure to the IMC, suggesting that its inclusion did not encourage participants to foil the study. That said, under different contexts with different established norms, participants could conceivably take offense at the inclusion of an IMC, especially in cross cultural studies. It is therefore important that researchers including IMCs be aware of this possibility, and structure their studies accordingly.

It is also worth noting that there are other ways of ensuring that participants read instructions that may be more effective than an IMC. For example, orally presenting the materials, or close supervision to increase motivation levels, could yield even higher quality data than the inclusion of an IMC. We encourage researchers to use such tools in the design of their studies to minimize the likelihood of satisficing. However, such approaches are not always possible, especially with increasingly popular web-survey studies. Moreover, even using such other methods, the addition of an IMC can be helpful. For example, in the preliminary studies reported in the introduction, participants were half as likely to fail an IMC in the presence of a supervisor, but 14% still failed. Thus, an IMC in conjunction with other methods for increasing participant diligence can help identify non-diligent participants if those other methods are not 100% effective.

We recommend using IMCs early in a study to convert satisficing participants into diligent participants, as in Study 2. This approach has the advantage that data from participants who fail the IMC are not excluded, thus preventing a reduction in sample

size in addition to reducing noise from non-diligent participants. However, such an approach may not always be possible (e.g. pen and paper studies) in which case eliminating satisficing participants may be necessary.

The noise created by participants who fail to read instructions decreases the reliability of the data and increases the expense associated with running studies as the number of participants necessary to achieve a reliable result is artificially increased. While the IMC may not be appropriate for every study, it is nonetheless a useful tool that can increase the efficiency of empirical research in psychology.

Acknowledgments

The authors would like to thank Jeff Zemla, Donna Shestowsky, Steven Miller, Sara Etchison and the Oppenheimer lab for feedback and assistance.

References

Arkes, H., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 124–140.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991–1013.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.

Glucksberg, S., & Haught, C. (2006). Can Florida become like the next Florida? When metaphoric comparisons fail. *Psychological Science, 17*, 935–938.

Kelley, H. (1950). The warm–cold variable in first impressions of persons. *Journal of Personality, 18*, 431–440.

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

Krosnick, J. A., Narayan, S. S., & Smith, W. R. (Eds.). (1996). *Satisficing in surveys initial evidence* Braverman, M. T., & Slater, J. K. (Eds.). *Advances in survey research* (pp. 29–44). San Francisco: Jossey-Bass.

Lee, A., & Aaker, J. (2003). Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion. *Journal of Personality and Social Psychology, 86*, 205–218.

Liberman, V., Samuels, S., & Ross, L. (2004). The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin, 30*, 1175–1185.

Loftus, E., & Palmer, J. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*, 33–43.

Page, S., & Bennesch, S. (1993). Gender and reporting differences in measures of depression. *Canadian Journal of Behavioural Science, 25*, 579–589.

Salancik, G. R., & Conway, M. (1975). Attitude inferences from salient and relevant cognitive content about behavior. *Journal of Personality and Social Psychology, 3*, 829–840.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology, 83*, 1178–1197.

Simon, H. A. (1957). *Models of man: Social and rational*. New York: John Wiley and Sons Inc..

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1*, 39–60.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199–214.