It’s a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on “Tricky” Tasks

David J. Hauser1 and Norbert Schwarz2

Abstract
Instructional manipulation checks (IMCs) have become popular tools for identifying inattentive participants in online studies. IMCs function by attempting to trick inattentive participants into responding incorrectly. However, from a conversational perspective, question characteristics are part of the researcher’s contribution to the conversation, and IMCs may teach participants that there is “more than meets the eye,” prompting systematic thinking on subsequent tricky-seeming questions in an attempt to avoid being tricked. In two online studies, participants responded to a simple task either before or after completing an IMC. As expected, answering an IMC prior to the task improved performance on items that benefit from increased systematic thinking—namely, the Cognitive Reflection Test (Study 1), and a probabilistic reasoning task (Study 2). We conclude that IMCs change attention rather than merely measure attention and discuss implications for their use in online studies.

Keywords
instructional manipulation checks, social cognition, systematic thinking, research methods

Instructional manipulation checks (IMCs) were designed to address the issue of inattentive responding in surveys and psychological tasks. Inattentive participants show increased context effects when answering questions, often responding with the first plausible response that comes to mind (Krosnick, 1999). This can contribute considerable noise to surveys and psychological studies. To identify inattentive respondents, Oppenheimer and colleagues introduced the IMC as a methodological tool. Although IMCs, also referred to as “trap” questions or attention checks, vary in surface characteristics, they mostly follow the same pattern: A lure question with lure responses is presented with an additional block of text that instructs participants to ignore the lures and to submit a non-intuitive response that they read in the instructions. Participants who select a lure response are considered inattentive. These “trapped” inattentive participants can contribute considerable error to data; hence, excluding them from data sets often assists with identifying reliable effects (Goodman, Cryder, & Cheema, 2013; Oppenheimer, Meyvis, & Davidenko, 2009).

As the amount of research conducted online has grown—from online surveys with Amazon Mechanical Turk (MTurk) workers to unsupervised collegiate subject pool participants potentially multitasking from the comfort of their dorms—concern about inattentive participants has increased. Many researchers suggest including IMCs in all online studies to identify inattentive participants (Goodman et al., 2013; Oppenheimer et al., 2009; Paolacci, Chandler, & Ipeirotis, 2010; Reis & Gosling, 2010), and some have even suggested including multiple IMCs, dispersed throughout the study, to measure variation in attention across the study (Berinsky, Margolis, & Sances, 2014). Indeed, IMCs have become popular as measures of attention and are now almost standard procedure in online research (Goodman et al., 2013; Peer, Vosgerau, & Acquisti, 2014).

However, IMCs may be more than a mere measure of attention—they may instead act as interventions that change how participants approach later questions. As research into the psychology of self-report illustrates, participants often infer a researcher’s intentions from aspects of the questionnaire, including the questionnaire’s lay-out, response format, and question sequence (for reviews, see Clark & Schober, 1992; Schwarz, 1994, 1999). Compared with many other questionnaire variables that have been found to influence response behavior, the message conveyed by IMCs is clear and salient: The researcher wants to know whether participants are paying attention. This focuses participants on an

1University of Michigan, Ann Arbor, USA
2University of Southern California, Los Angeles, USA

Corresponding Author:
David J. Hauser, Department of Psychology, University of Michigan, 3233 East Hall, 530 Church St., Ann Arbor, MI 48109-1043, USA.
Email: dhauser@umich.edu
aspect they may otherwise not have in mind—Paying close attention is presumably important for this task and apparently highly valued.

IMCs also attempt to lure participants into responding incorrectly. By containing lure questions and responses, IMCs ask participants to override their original response, and instead input an answer that would often be uncooperative and uninformative in a standard question (Berinsky et al., 2014; Goodman et al., 2013; Oppenheimer et al., 2009). Thus, IMCs may also inform participants that the researcher is an uncooperative communicator who may try to lure them into responding incorrectly. This highlights that researchers may craft questions that contain more than meets the eye. Thus, when encountering a later task that may seem a bit tricky, participants may assume that there are non-obvious aspects to the task and may adopt more systematic thinking in an attempt to avoid being tricked again.

These considerations suggest that IMCs can change how much attention participants pay to the task, thus changing their response behavior. We test this possibility in two studies that present tasks on which increased systematic thinking should improve performance.

**Study 1: Trap Questions Improve Performance on the Cognitive Reflection Test (CRT)**

The CRT (Frederick, 2005) is an individual difference measure based on three numeric reasoning problems. For each problem, an intuitively appealing answer is likely to come to mind quickly but is incorrect. Instead, arriving at the correct answer requires applying the appropriate formula, performing calculations, and checking one’s solution. The CRT is assumed to be a measure of a person’s stable disposition to engage in reflective, systematic thought, predicting performance on heuristics-and-biases tasks (Toplak, West, & Stanovich, 2011), delay of gratification, and Wundertlic IQ scores (Frederick, 2005). Furthermore, numerous studies have shown that task order does not affect CRT scores. For instance, CRT scores are similar regardless of whether the CRT comes before or after responding to various moral dilemmas (Paxton, Bruni, & Greene, 2014; Paxton, Ungar, & Greene, 2012; Pinillos, Smith, Nair, Marchetto, & Mun, 2011). Thus, it is often assumed that contextual cues in a survey have no impact on CRT scores. If the IMC acts as a mere measure of attention, then CRT scores should be similar regardless of whether the IMC occurs before or after the CRT.

However, note that correctly answering CRT problems requires checking and overriding an intuitive answer, much as passing an IMC requires checking the instructions and overriding the lure response. If the IMC teaches participants that there is more than meets the eye to the questions, then varying the order of the IMC and the CRT should affect participants’ CRT scores. We therefore predicted that participants who receive an IMC question prior to the CRT will score better on the CRT than participants who complete the tasks in the reverse order. If so, this effect would be especially powerful considering that order effects on CRT scores are unobserved in the literature on the CRT (Paxton et al., 2014; Paxton et al., 2012; Pinillos et al., 2011).

**Method**

**Participants.** Three hundred eighty workers (224 male, 154 female, 2 unidentified) from MTurk participated in the study in exchange for 50 cents.

**Materials and procedure.** Participants were directed to a survey on decision making, administered via Qualtrics. All participants completed two tasks: the IMC and the CRT. Based on random assignment, some participants received the IMC prior to the CRT, while others received the CRT prior to the IMC.

The IMC was adapted from Oppenheimer et al. (2009). Under the header “Sports Participation” respondents were asked, “Which of these activities do you engage in regularly? (click on all that apply).” However, above the question, a block of instructions (in smaller text) indicated that, to demonstrate attention, respondents should click the other option and enter “I read the instructions” in the corresponding text box. Following these instructions was scored as a correct response to the IMC.

Participants were also randomly assigned to receive feedback on their IMC response or not. If their initial IMC response was incorrect, those receiving feedback were returned back to the IMC question with the instructions “Please try again.” Those not receiving feedback simply progressed to the next part of the task regardless of IMC accuracy. However, the feedback manipulation did not have any main effects, nor did it moderate any of the current results. It will therefore not be discussed further.

The CRT task, adapted from Frederick (2005), asked participants three numeric questions that necessitated the suppression of an intuitive, but incorrect, response to answer correctly. Text entry boxes were used to record participants’ responses. The questions, intuitive but incorrect answers in parentheses, and actual answers, follow:

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (10) 5 cents
If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (100) 5 minutes
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (24) 47 days

After completing the CRT and IMC, we probed whether participants were familiar with the IMC and the CRT.
Participants were asked whether they’ve seen the “Sports Participation” question before (IMC) and the “bat ball” question before (CRT), responding yes or no.

**Results and Discussion**

Three hundred fifty participants (92%) responded correctly to the IMC on their first try, while 30 participants did not. This IMC pass rate is consistent with those of recent research on MTurk (Nauts, Langner, Huijsmans, Vonk, & Wigboldus, 2014; Wolf et al., 2014; for related research, see Hauser & Schwarz, in press). Passing the IMC was not affected by IMC presentation order, \( \chi^2 (1, N = 380) = 0.98, p = .321 \). Therefore, we limited our analysis to those 350 participants who passed the IMC because the group of participants who responded incorrectly was not large enough to draw firm conclusions.1 Furthermore, studies that use IMCs typically include only the participants who respond correctly to the IMC, and answering the IMC should only affect participants who recognize that the question is intended to trick inattentive participants (i.e., those who answered the IMC correctly). Thus, this group is the most theoretically relevant population for the current analyses.

Overall, CRT scores were comparable with those observed in prior research (West, Meserve, & Stanovich, 2012, Study 2). More important, IMC presentation order significantly affected CRT scores as shown in the first row of Table 1. Answering the IMC first raised CRT scores relative to answering the CRT first. This effect was neither moderated by participants having seen the IMC in prior MTurk studies (two-way interaction of IMC order and prior exposure to the IMC: \( F < 1 \)) nor moderated by participants having seen the CRT in prior MTurk studies (two-way interaction of IMC order and prior exposure to the CRT: \( F < 1 \)).

IMC presentation order also significantly affected intuitive responses to the CRT problems and the number of seconds spent answering the CRT questions. As shown in the second and third rows of Table 1, answering the IMC first lowered the number of intuitive responses and increased the number of seconds (log-transformed) participants spent completing the problems (relative to answering the CRT first).

In sum, prior exposure to an IMC increased CRT scores and increased the time spent working on the CRT. These effects were not moderated by familiarity with the CRT or the IMC. As the CRT is an individual difference measure with links to IQ (Frederick, 2005) and which is rarely affected by task order (Paxton et al., 2014; Paxton et al., 2012; Pinillos et al., 2011), IMCs appear to be powerful manipulations of attention and systematic thinking, highlighting that they operate as interventions rather than mere measures.

### Study 2: Trap Questions Improve Performance in a Probabilistic Reasoning Task

To further test the impact of IMCs on cognitive performance, Study 2 employs a different task, namely, a probabilistic reasoning task (described below) on which performance is impaired by denominator neglect (Toplak et al., 2011). As observed in earlier research, probabilistic reasoning is often improved by systematic thinking (e.g., Denes-Raj & Epstein, 1994; Toplak et al., 2011). Therefore, if exposure to an IMC promotes systematic thinking, it should improve performance on a probabilistic reasoning task.

#### Method

**Participants.** Four hundred two workers (220 female, age range 18-73) from MTurk completed the task in exchange for 20 cents each.

**Materials and procedure.** Participants were directed to a survey on decision making, where they completed three questions. All participants first received a filler question about their personality traits that mimicked the formatting of Oppenheimer et al.’s (2009) IMC. This question was included to make the IMC appear consistent in formatting with prior questions. This filler task also ensured that any order effect observed on the probabilistic reasoning task could not be attributable to the presence of any non-specific prior task. That is, because both question orders received the filler task first, any advantage of the IMC first condition must be attributable to receiving, specifically, the IMC question first (rather than receiving any task first, in general).

Following the filler question, all participants received an IMC and the reasoning task in a randomly assigned order:

| IMC presentation order | Measure                | IMC first | IMC last | F   | d    |
|------------------------|------------------------|-----------|----------|-----|------|
| Correct CRT score      | 1.76 (1.16) [1.59, 1.93] | 1.33 (1.10) [1.17, 1.49] | 12.64*** | .38 |
| Intuitive CRT score    | 1.03 (1.10) [0.87, 1.19] | 1.37 (1.06) [1.21, 1.53] | 8.59**   | .31 |
| Time spent (in log-transformed seconds) | 3.65 (0.84) [3.52, 3.78] | 3.48 (0.77) [3.36, 3.59] | 3.99*    | .21 |

Note. Standard deviations in parentheses; 95% confidence intervals in brackets. IMC = instructional manipulation check; CRT = Cognitive Reflection Test. \* \( p < .05 \). \** \( p < .01 \). \*** \( p < .001 \).
Half of the participants received the IMC first (followed by the reasoning task on the next page), while the other participants received the reasoning task first (followed by the IMC on the next page).

The IMC (adapted from Oppenheimer et al., 2009) was identical to the sports participation IMC used in Study 1. Because feedback on the IMC did not affect results in Study 1, the IMC was administered without feedback in Study 2.

The probabilistic reasoning task (adapted from Toplak et al., 2011) presented participants with a hypothetical scenario in which they were to draw a black marble from one of two trays (without looking) to win a prize. There was a small and large tray, and participants were tasked with choosing the tray that would increase their odds of winning. The small tray contained 1 black marble and 9 white marbles, whereas the large tray contained 8 black and 92 white marbles. Participants typed their preferred tray into a text box. Objectively, the small tray offers a greater likelihood of drawing a black marble (1/9) than the large tray (8/92), but people often neglect the denominator and prefer the tray with the larger number of black marbles. Accordingly, choosing the small tray was scored as the analytic and choosing the large tray as the intuitive response.

Following these questions, we probed whether participants had previously been exposed to the “Sports Participation” attention check question and the “which tray would you draw a marble from” question (yes/no).

Results and Discussion

Consistent with prior research (Hauser & Schwarz, in press; Nauts et al., 2014; Wolf et al., 2014) and Study 1, most participants (89%) responded correctly to the IMC. Moreover, passing the IMC was not affected by IMC order (89% in either order). We limit the analysis to the 358 participants who passed the IMC. Note that excluding inattentive participants is the purpose for which IMCs are included in online studies, so dropping the 44 participants who failed the IMC is consistent with common practice and the intended use of IMCs.

We hypothesized that when the IMC came first, it would increase optimal responding on the denominator neglect question. This was the case. More participants selected the optimal response on the denominator neglect question when the IMC preceded it (75% correct) than when the IMC followed it (62% correct), $\chi^2 (1, N = 358) = 7.41, p = .006, \varphi = .14$.

In addition, the effect of IMC order was not moderated by familiarity with the IMC or familiarity with the denominator neglect question. We conducted a logistic regression entering familiarity with the IMC (no vs. yes), familiarity with the denominator neglect question (no vs. yes), IMC order (first vs. last), and their interactions as mean-centered predictors of denominator neglect responses (non-optimal vs. optimal). The main effect of IMC order was again significant, $B = -.63, Wald = 7.24, p = .007$, odds ratio $= .53$, and was neither moderated by familiarity with the IMC (two-way interaction of IMC familiarity × IMC order, $B = -.37, Wald = 0.63, p = .43$) nor moderated by familiarity with the denominator neglect question (two-way interaction of denominator neglect familiarity × IMC order, $B = .28, Wald = 0.14, p = .72$). All other effects in the logistic regression model failed to reach significance, $ps > .37$.

In sum, prior exposure to an IMC improved performance on a probabilistic reasoning task. These effects were not moderated by familiarity with the task or the IMC. As non-optimal responding in this task predicts real-life risky behavior (such as gambling; Denes-Raj & Epstein, 1994), this study echoes the conclusions of Study 1: IMCs are more than mere measures of attention; they are interventions that alter attention and systematic thinking on subsequent tasks.

General Discussion

Our findings converge to show that an IMC is more than merely a measure of attention—It is an intervention that increases systematic thinking on subsequent tricky-seeming tasks. When highly accessible and intuitively appealing answers are incorrect, a trap question can improve performance by prompting participants to reconsider their spontaneous solution and adopt a more systematic reasoning strategy. In Study 1, this resulted in improved performance on the CRT (Frederick, 2005), an individual difference measure that is associated with rational problem solving. In Study 2, this resulted in increased optimal responses on a probabilistic reasoning task also associated with rational problem solving (Denes-Raj & Epstein, 1994; Toplak et al., 2011). In combination, these findings show that an IMC can signal a context in which questions may contain “more than meets the eye,” a conclusion that significantly alters participant’s reasoning strategies compared with participants not exposed to an IMC.

These studies have important implications for the use of IMCs in online tasks. IMCs are typically conceptualized as measures, not interventions. However, as demonstrated here, an IMC prompts more systematic reasoning on subsequent tricky-seeming tasks. Accordingly, IMC administration could have moderating effects on a study’s results. For instance, systematic thinking could attenuate or eliminate effects expected to be driven by intuitive reasoning and could amplify effects expected to be driven by reflective reasoning. One should therefore exercise caution in IMC use, as IMCs could eliminate effects one might otherwise observe or create effects that might otherwise not emerge. We assume that these complications are most likely under conditions where intuitive and systematic reasoning can lead to different answers, but future research may fruitfully explore whether the influence of IMCs extends beyond this type of task. In addition, different types of IMCs may cause different effects. For instance, more novel or more difficult IMCs may more strongly prompt systematic thinking, whereas commonplace IMCs may have little influence. The observed effects may also depend upon whether participants have the knowledge...
or skills required to complete the systematic thinking task. For example, an IMC may send the message to be on the lookout for trick questions, but if a participant does not have the mathematical skills to know that their intuitive answer of 10 cents to the bat ball problem is incorrect, then a prior IMC may do little to increase their accuracy on such tasks. These are interesting questions, which we will leave for future research to explore.

Our results implore a seemingly obvious suggestion: If IMC use is necessary, place the IMC after all crucial measures and manipulations in a study. Reserving IMCs for the end of a survey at least guarantees that the IMC will not affect how participants interpret and answer other questions. However, placing IMCs at the end does have drawbacks. Inattentiveness may not be merely a trait, measureable with only one IMC, but instead may uniquely ebb and flow for each participant throughout the tasks (Berinsky et al., 2014). Further, participants often become fatigued as they progress through long surveys, and survey fatigue predicts satisficing behaviors (Krosnick, 1999), which are expected to also positively predict IMC failure (Oppenheimer et al., 2009). For these reasons, an IMC placed at the end of a survey may not accurately assess a participant’s attentiveness to the previously administered questions and thus may misidentify who did or did not pay attention at an earlier point.

Acknowledgments
We thank Aashna Sunderrajan and Madhuri Natarajan for their assistance in conducting the studies.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research and/or authorship of this article.

Notes
1. Because feedback was only (randomly) given to participants who failed the instructional manipulation check (IMC), participants who received feedback were not included in our subsequent comparisons.
2. In all, 76.3% of participants indicated having seen the Cognitive Reflection Test (CRT) problems before and 44.6% indicated having seen the IMC before. This is unsurprising as these tasks have become quite popular on Amazon Mechanical Turk (MTurk) in recent years (Chandler, Mueller, & Paolacci, 2014).

References
Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. American Journal of Political Science, 58, 739-753.
Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. Behavior Research Methods, 46, 112-130.
Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), Questions about questions: Inquiries into the cognitive bases of surveys (pp. 15-48). New York, NY: Russell Sage.
Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. Journal of Personality and Social Psychology, 66, 819-829.
Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives, 19, 25-42.
Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. Journal of Behavioral Decision Making, 26, 213-224.
Hauser, D. J., & Schwarz, N. (in press). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behavior Research Methods. Advance online publication. doi:10.3758/s13428-015-0758-z.
Krosnick, J. A. (1999). Survey research. Annual Review of Psychology, 50, 537-567.
Nauts, S., Langner, O., Huijsmans, I., Vonk, R., & Wigboldus, D. H. (2014). Forming impressions of personality. Social Psychology, 45, 153-163.
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.
Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. Judgment and Decision Making, 5, 411-419.
Paxton, J. M., Bruni, T., & Greene, J. D. (2014). Are “counter-intuitive” deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). Social Cognitive and Affective Neuroscience, 9, 1368-1371.
Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. Cognitive Science, 36, 163-177.
Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behavior Research Methods, 46, 1023-1031.
Pinillos, N. A., Smith, N., Nair, G. S., Marchetto, P., & Mun, C. (2011). Philosophy’s new challenge: Experiments and intentional action. Mind & Language, 26, 115-139.
Reis, H. T., & Gosling, S. D. (2010). Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), Handbook of social psychology (5th ed., Vol. 1, pp. 82-114). New York, NY: John Wiley.
Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), Advances in experimental social psychology (Vol. 26, pp. 123-162). San Diego, CA: Academic Press.
Schwarz, N. (1999). Self-reports: How the questions shape the answers. American Psychologist, 54, 93-105.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory & Cognition, 39, 1275-1289.
West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. Journal of Personality and Social Psychology, 103, 506-519.
Wolf, W., Levordashka, A., Ruff, J. R., Kraaijeveld, S., Lueckmann, J. M., & Williams, K. D. (2014). Ostracism Online: A social media ostracism paradigm. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-014-0475-x

**Author Biographies**

**David J. Hauser** is a doctoral student in social psychology at the University of Michigan. His research focuses on the role of language processes, including conceptual metaphors, on thought and reasoning, and on cognitive aspects of survey methods.

**Norbert Schwarz** is provost professor of Psychology and Marketing, and founding co-director of the Dornsife Mind and Society Center, at the University of Southern California. His research focuses on human judgment and decision making, particularly the socially situated and embodied nature of cognition, the interplay of feeling and thinking, and the implications of basic cognitive and communicative processes for public opinion, consumer behavior, and social science research.