

SHORT RESEARCH NOTE

On the Limitations of Manipulation Checks: An Obstacle Toward Cumulative Science

Marie-Pierre Fayant*, Harold Sigall†, Aurore Lemonnier*, Emilie Retsin* and Theodore Alexopoulos*

Manipulation checks do not allow ruling out or accepting alternative explanations of causal effects (Sigall & Mills, 1998). In order to gauge the influence of this argument on current research practices, we surveyed the views of researchers on manipulation checks. Results confirmed that a manipulation check still stands as a totem of experimental rigor. Except in rare circumstances, such as when pilot testing, manipulation checks do not provide information relevant to construct validity. While it seems cost free to include seemingly informative manipulation checks, we claim it is actually costly because it wrongly enhances subjective confidence in the validity of research findings. We conclude that manipulation checks may hinder efforts to adopt a cumulative culture and practice of hypothesis testing.

Keywords: best research practices; construct validity; experimentation; manipulation check

Experimentation represents a key method for inquiring into the determinants of social behavior. A social psychology experiment typically involves manipulating an element of the situation to observe its effect on another element. Often, researchers include measures of the manipulation to assess its effectiveness. The current paper discusses the necessity and function of these measures in view of recent methodological concerns. We argue that despite their widespread and established use, measures of manipulations are neither necessary, nor generally useful and may actually impede scientific progress.

On validity and manipulation checks

The goal of social psychology consists of *explaining* social phenomena. Researchers are concerned with demonstrating causal relationships as well as understanding why, how and under what conditions such causal relationships occur (Brewer & Crano, 2014). Drawing causal conclusions requires demonstrating that the manipulated variable actually causes some measured variations (*internal validity*), and that the causal link between concrete operations can be generalized to relevant theoretical concepts (*construct validity*; Brewer & Crano, 2014; Shadish, Cook, & Campbell, 2002). To ensure that the manipulated independent variable (IV) is a valid instantiation of the

conceptual variable, researchers often include measures of experimental manipulations termed *manipulation checks* (MCs).¹

In a seminal paper, Sigall and Mills (1998) argued that MCs were not necessary to establish *construct validity of causes and effects*, that is, to reach the conclusion that operational manipulations and measurements are unambiguously linked at the conceptual level. According to the authors, whether or not alternative explanations for the observed effects exist represents the essential grid of analysis for the (potential) added value of MCs. In its core, their rationale is as follows: *when no alternative explanations exist*, a successful MC does not constitute additional proof for construct validity, and a failed MC does not invalidate the theoretically expected empirical result. *When alternative explanations exist*, a successful MC in no way rules out other plausible accounts and a failed MC does not constitute definitive evidence against the favored explanation. Interestingly, Sigall and Mills mentioned that when the manipulation does not produce the intended effect, a positive MC may rule out the possibility that the treatment was unsuccessful in varying the conceptual IV. They nevertheless emphasized that this information could come from other sources like secondary dependent variables (DVs). Finally, they extended their argument to the class of mediating variables as: “From an experimental point of view, a mediator check is similar to an independent variable check” (p. 225). The paper concluded that the inclusion of either manipulation or mediator checks fails to provide definitive information for ruling out (or accepting) alternative explanations of a causal effect.

Although the argument seems to have been influential for mediation (Fiedler, Schott, & Meiser, 2011; Jacoby &

* Laboratoire de Psychologie Sociale: Menaces et Société (EA 4471), Université Paris Descartes, FR

† Department of Psychology, University of Maryland, College Park, US

Corresponding authors: Marie-Pierre Fayant (marie-pierre.fayant@parisdescartes.fr) and Theodore Alexopoulos (theodore.alexopoulos@parisdescartes.fr)

Sassenberg, 2011), it is not so for the use of MCs. According to Haslam and McGarty (2003), MCs are almost a mandatory requirement for research reports to survive the reviewing process. In our opinion, publications that provide guidance on methodological matters do not sufficiently warn against the non-informative value of MCs for construct validity (but see O'Keefe, 2003, for an exception). They either fail to discuss MCs (e.g., Reis & Judd, 2014) or mention their potentially informative function—for instance when results contradict the predictions—while remaining silent about other cases (Wilson, Aronson, & Carlsmith, 2010). Some even advocate their use by considering that a MC is an essential element for asserting internal and construct validity of an experiment (Flake, Pek, & Hehman, 2017; Foschi, 2014). More recently, MCs have been presented as necessary elements in *close replications* (Hüffmeier, Mazei, & Schultze, 2016; Stroebe & Strack, 2014).

Assessing the views of social psychologists: a field experiment

In order to get a glimpse of the progress of Sigall and Mills' argument among scholars, we explored current beliefs regarding MCs in relation to construct validity. We surveyed 101 researchers (among a total of 198) attending the 2016 Geneva meeting of the Association for the Diffusion of International Research in Social Psychology. To do so, we tested the impact of the presence of a MC in an experimental design. Then, following Sigall and Mills, we assessed the general views regarding MCs.

Respondents were asked to role-play reviewers evaluating a paper submitted to a conference. They read an abstract of an experiment examining the impact of heuristic cues (communicator's likeability) on students' attitudes in a low-involvement setting—a conceptual replication of Chaiken (1980). The operationalization of communicator's likeability was the extent to which she declared her commitment to student-related activities (she declared to be fully vs. lightly committed). In such a case, alternative explanations clearly exist (e.g., communicator's perceived status, participants' mood). We decided to use a scenario in which alternative explanations exist because we wanted to assess situations where construct validity is at stake (Brewer & Crano, 2014). The abstract specified that the message was in favor of work-time arrangement for public workers. The DV was agreement with the message content on a 7-point scale (1: very unfavorable; 7: very favorable). The relationship between the IV (communicator's likeability) and the DV was presented as statistically significant.

All respondents received the same abstract; however, half of them read that the experiment included a successful MC (communicator's likeability rating) whereas no MC was mentioned for their counterparts. Then participants rated their confidence in the data (Items 2 and 5; Items 1, 3 and 4 were fillers) as well as the necessity of the inclusion of MCs in a well-designed experiment (Items 6 to 9 derived from Sigall and Mills, see **Table 1**). Regarding the items specifically designed to assess confidence in the data, Item 2 asked participants to indicate their certainty that the source's sympathy created a more favorable evaluation of the message content (1: not at all

certain; 10: completely certain). Item 5 asked whether the addition/presence of a MC allows the conclusion that the source's sympathy created a more favorable evaluation of the message content (Yes/No).

Results (see **Table 1**) show that a MC still stands as a totem of experimental rigor as the confidence that the IV caused the observed changes in the DV was lower in the MC absent ($M = 4.53$, $SD = 2.02$) than MC present condition ($M = 5.12$, $SD = 2.03$), $t(99) = 1.45$, $p = 0.15$, $d = 0.29$, 95% CI $[-0.10, 0.68]$. Although not statistically significant, the descriptive means are in line with the idea that the MC influenced confidence in our sample.

Corroborating this result, the MC's perceived value (Item 5) was greater under the MC absent than MC present condition, $z = 2.94$, $Chi^2(1) = 8.7$, $p = 0.003$, $OR = 3.59$, 95% CI $[1.56, 8.52]$. In the MC absent condition, 71.43% respondents answered that adding a MC would allow them to reliably conclude that the communicator's likeability was the cause of persuasion. However, in the MC present condition, only 40.38% considered that the actual presence of such a measure allows a reliable conclusion.

Moving to the items taken from Sigall and Mills and pooling across experimental conditions, 78.26% (vs. 17.84%) of the respondents answered positively that a MC was necessary in a well-designed experiment (Item 6), a result that is above the 60% found by Sigall and Mills. Complementing this finding, the item assessing whether the absence of a MC constitutes a methodological flaw (Item 7) received more affirmative (55.61%) than negative (43.3%) answers (67% answered positively in the Sigall and Mills paper). Overall, this survey indicates that researchers—at least those who were attending this specific meeting—still value MCs for construct validity issues.

Benefits and costs of MCs

In support of MCs. A MC is customarily considered an informative tool (Foschi, 2014; Hüffmeier et al., 2016; Stroebe & Strack, 2014; Wilson et al., 2010). An experiment yielding evidence that (a) the IV has the intended effect on the DV, and (b) the experimental groups are contrasted in terms of the MC is taken as providing cogent evidence for the claimed causal relationship. For this reason, an MC is considered by some as informative regarding internal and construct validity (Flake et al., 2017). In some specific cases where the IV produces an effect on the DV but fails to affect the MC, some researchers may take this as useful evidence in favor of an alternative explanation.

An influential argument for using MCs is that their informative value may be substantial when the IV does *not* produce the intended effect on the DV. In such cases, a successful MC could potentially rule out the possibility that the manipulation was not successful in varying the conceptual variable (Sigall & Mills, 1998), and may suggest either that the treatment was not strong enough to produce variations on the DV (Haslam & McGarty, 2003) or that the hypothesis was wrong (Wilson et al., 2010). Accordingly, they are mentioned in best practice recommendations as tools to gain information when conducting (close) replications (Hüffmeier et al., 2016; Stroebe & Strack, 2014).

Table 1: Results as a function of the presence versus absence of a manipulation check.

| Questions ^a | Without MC (<i>n</i> = 49) | | With MC (<i>n</i> = 51) | |
|--|------------------------------|----------------|---------------------------|----------------|
| | <i>M</i> (<i>SD</i>) | | <i>M</i> (<i>SD</i>) | |
| 1. With which degree of certainty would you say that the sample's size is sufficient to test the researcher's hypothesis? | 5.84 (2.37) | | 6.92 (2.02) | |
| 2. With which degree of certainty would you say that the source's sympathy created a more favorable evaluation of the message's content? | 4.53 (2.06) | | 5.12 (2.03) | |
| Yes/No questions | Without: MC (<i>n</i> = 49) | | With: MC (<i>n</i> = 51) | |
| | Answer Yes % | Answer No % | Answer Yes % | Answer No % |
| 3. Does this study contain all the methodological precautions to conclude that the source's sympathy created a more favorable evaluation of the message content? | 14.29 | 83.67 | 21.15 | 76.92 |
| 4. Would you accept this communication in a congress' program? | 85.71 | 14.29 | 80.77 | 17.31 |
| 5. Does the addition/presence of a MC allow concluding that the source's sympathy created a more favorable evaluation of the message content? | 71.43 | 26.53 | 40.38 | 53.85 |
| 6. In your opinion, is a MC necessary in a well-designed social psychology lab experiment? | 79.59 | 18.37 | 76.92 | 17.31 |
| 7. In your opinion, would the absence of the MC constitute a flaw? | 61.22 | 38.78 | 50.00 | 48.08 |
| 8. In your opinion, is a MC necessary for the causality relation between an IV and a DV? | 51.02 | 48.98 | 40.38 | 59.62 |
| 9. In your opinion, if a researcher wants to test a mediation hypothesis, is a measure of the mediator between the IV and the DV necessary? | 77.55 | 16.33 | 75.00 | 21.15 |

^a Ratings on 10-point scales (1: not at all certain; 10: completely certain). Items appear in the order of presentation. Here we report estimations per condition for filler items (1, 3, and 4) and items assessing general view of MC (6–9).

The perils of MC. In order to expose the misuses of MCs, it is important to highlight the distinction between internal validity and construct validity of causes and effects. Internal validity refers to the extent to which one is confident that the manipulated IV created the observed variations in a particular experiment. Construct validity concerns the generalization and the inferences of this causal link to some relevant theoretical concepts. Thus, problems of internal validity generally arise from experimental flaws (e.g., self-selection, experimental artefacts), whereas construct validity issues appear when some potential theoretical confounds plague the explanation of the results (Brewer & Crano, 2014). That being said, it is important to note that an MC cannot provide evidence for causality as it is mute regarding internal validity. For instance, one might show the effectiveness of an intervention while disregarding the theoretical reasons of such an effect (an example of Sigall and Mills' *no alternative explanation* case). As mentioned above, even in such a situation a MC is irrelevant as it cannot validate or invalidate the (very) fact that the only altered element is the intervention. Actually, regarding

strictly causal relationship issues, (un)successful MCs are uninformative.

Regarding construct validity of causes and effects—whether the focal conceptual IV is implicated in the observed causal variation—the MC is also limited as it is not a definitive empirical shield against alternative explanations (Sigall & Mills, 1998). When the results are positive, the manipulation may have affected different constructs among which the conceptual IV (supposedly measured by the MC) represents only one instance. In that case, it is impossible to know which construct affected the DV, and a successful MC cannot resolve this ambiguity. Hence, a positive MC cannot sustain the focal hypothesis. In a related vein and contrary to what has been discussed above, a failed MC does not speak in favor of an alternative explanation for two main reasons: (a) it could be attributed to measurement problems and (b) such a measure is not designed to assess the viability of alternative explanations.² Consequently, any attempt to use information stemming from a failed MC in favor of other explanations is unwarranted. In spite of this

reasoning, Sigall and Mills still argued that MCs could be informative in a case of a failed experiment. A successful MC could tell us that the conceptual IV was successfully manipulated and may suggest that the hypothesis was wrong. Yet, such a systematic variation observed on an “informative” MC could actually be due to some covariate of the conceptual variable. Therefore, concluding that the manipulation is valid would be unwarranted. Relatedly, because of this concern researchers cannot rely on positive MCs to reach conclusions in terms of (the lack of) manipulation strength.

Aside from validity issues, the MC presents several well-known shortcomings (Bless & Burger, 2016; Kühnen, 2010). Its inclusion may lead to unpredicted results because it could render salient the manipulation, redirect attention to the research goal, and lead to counter- or overcorrection attempts for the manipulation’s perceived influence. Conversely, a MC may well create the predicted effect, either through experimental demand or by setting in motion a psychological process. More generally, where to place the MC in an experiment is always a puzzle to the experimenter: placed before the DV it can be a source of contamination, whereas placed after one runs the risk of obtaining null effects because the treatment impact might have dissipated (e.g., affective states).

Finally, MC consists of adding a measure to the experiment. Conducting multiple tests increases Type I error rate (Cohen, 1990) and endangers conclusions drawn from the results. As experiments containing multiple DVs have less chance to show significant results on every measure than on any one of them (Maxwell, 2004), an MC decreases the power to observe statistically significant results on every measure while such effects indeed exist (increasing the risk of committing a Type II error).

To summarize, MCs are uninformative about internal and construct validity. Moreover, by corrupting the process under study, its inclusion could thwart internal validity. MCs could also endanger conclusions drawn from observed results by increasing Type I and Type II error rates. Although researchers could sometimes be inclined to take this risk when multiplying the number of measures, we believe this risk is not worth taking in the case of MC, given its costs. Despite this, some authors would still argue that, in cases of non-predicted results, a successful MC provides some information (Sigall & Mills, 1998; Wilson et al., 2010). Although we take note of this position, we nevertheless believe it generally represents a relatively small benefit.

About construct validity

Construct validity reflects an evaluative judgment on the fit between theoretical and empirical arguments and the interpretation that the operationalization is an appropriate translation of the concept (Messick, 1995). As reaching a conclusion on validity is a subjective process, one needs to accumulate a good deal of arguments to constrain any potential decision biases. Although an MC seems to be a handy recipe to ascertain validity, it is not a well-suited instrument for this goal. Tackling construct validity issues requires a rigorous scientific posture and the use of stringent procedures akin to theory testing (Brewer & Crano,

2014; Messick, 1995; Shadish et al., 2002). The validation process requires an accumulation of evidence including: (i) a theoretical evaluation of the translation of the concept into its implementation; (ii) convergence and discrimination demonstrations based on an empirical set of correlations with related and unrelated constructs respectively (e.g., tests of moderation); and (iii) the prediction of external criteria such as new DVs.

Obviously, the first basic ingredient needed for high construct validity is a comprehensive theoretical framework. Only then can the researcher achieve a rigorous and systematic description of the phenomenon under study. As theoretical concepts represent abstract verbal definitions that need to be translated into their referents in the real world (Deutsch & Krauss, 1965), the more precise and exhaustive their definition is, the more unambiguously the concepts are tailored into discrete and meaningful operations (Cook, Campbell, & Perrachio, 1990). As each operational translation may include a unique part of noise or irrelevancy and/or omit theoretically pertinent components, experimental manipulations are rarely a perfect instantiation of the conceptual variable. A classical recommendation would be to rely on *multiple operations* of the IV in order that the various treatments are associated with a diverse sample of irrelevant factors, so that they do not systematically covary with the focal variable (convergent and discriminant validity, Brewer & Crano, 2014, Cook et al., 1990; Lench, Taylor, & Bench, 2013). Theoretically valid conclusions are achieved through well-devised experimental research programs based on solid theoretical grounds that systematically address alternative hypotheses, what Platt (1964) coined *strong inference*. As such, they should always represent crucial tests that provide elements for the *exclusion* of a hypothesis. The resulting theoretical refinement is achieved through a repetitive sequence of uniquely useful experiments which conform to a conditional inductive process.

MC with respect to scientific practice and cumulative science

Given the complex process required to judge an operationalization as valid, MCs are by no means able to strengthen conclusions in terms of validity of causes and effects. Indeed, just as any other measure, MC is vulnerable to measurement issues (e.g., sensitivity, reliability) and requires construct validation before its inclusion in an experiment. As already argued, MCs do not warrant conclusions of causal validity and they actually present several methodological caveats. While at first glance MCs seem to be cost-free and to provide an informational benefit, including them can be costly on several levels. Crucially, we believe that MCs may wrongly enhance subjective confidence in the operationalization. By doing so, a successful MC may draw the researcher’s attention to a particular conceptual variable as an explanation for the effect whilst neglecting countless other variables that were not measured in the experimental design but that could still contribute to the phenomenon (see Fiedler et al., 2011). By increasing subjective confidence, we fear that MCs may lower the need to conduct extensive replications of the results through multiple operations and eventually lead

to a mono-operation bias. The inclusion of MCs may thus insidiously thwart efforts toward a cumulative culture and practice of hypothesis testing. In terms of a cost-benefit analysis, we therefore argue that relying on MCs is suboptimal for corroborating the validity of research findings.

Importantly, in the context of discussions on best research practices (Finkel, Eastwick, & Reis, 2015), MCs may represent an obstacle against cumulative knowledge culture simply because researchers may spend time pondering failed MCs or make unwarranted inferences from successful MCs. Our analysis seems at odds with recent recommendations for replicability with the inclusion of MCs (Hüffmeier et al., 2016; Stroebe & Strack, 2014). Such recommendations follow from construct validity concerns as well as issues related to the comparability of the operational definitions between the original and replication experiments (especially closed replications, Brandt et al., 2014). We concur with such propositions to the extent that the informational gain contributes to convincing replications, but believe that decisive information comes from other sources like secondary DVs, pretests, and pilot experiments (Wilson et al., 2010).

On another level, abandoning MCs would potentially relieve researchers from failures to report all included measures, a widespread hurdle to best practices (John, Loewenstein, & Prelec, 2012). This might even redirect them to pay extra care to the concrete IV and DV operationalizations. Such a positive practice should decrease the proportion of failed results reported in scientific communications. Also taking advantage of Internet open-access resources, one can make available data from pretests, pilots, and previous (failed) hypothesis tests instead of relying on MC. This view fits nicely with recent recommendations for best practices in social psychology.

Concluding remarks

Almost 20 years ago, Sigall and Mills highlighted that MCs were unnecessary, but their argument seemingly failed to reach its audience. This paper fills this gap in reaffirming the non-necessity of including MCs in experimental research and goes a step further in arguing that MC might work against cumulative practice. We hope this paper will serve as a call back to the fundamentals of experimentation with a strong emphasis on construct validity, and shift back researchers' attention toward theorization, replicability, and testing of logical alternatives.

Notes

- ¹ We refer to *manipulation checks* as measures tapping the conceptual variable and not checks of whether participants perceived the treatment or followed the instructions.
- ² Foschi (2004) recommends implementing additional MCs to assess alternative explanations. In our opinion, such recommendation is not viable as it is impossible to generate and assess each and every plausible alternative explanation.

Competing Interests

The authors have no competing interests to declare.

References

- Bless, H., & Burger, A. M.** (2016). A closer look at social psychologists' silver bullet: Inevitable and evitable side effects of the experimental approach. *Perspectives On Psychological Science, 11*, 296–308. DOI: <https://doi.org/10.1177/1745691615621278>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., van't Veer, A., et al.** (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*, 217–224. DOI: <https://doi.org/10.1016/j.jesp.2013.10.005>
- Brewer, M. B., & Crano, W. D.** (2014). Research design and issues of validity. In: Reis, H. T., & Judd, C. M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*, 2nd ed. (pp. 11–26). New York, NY, US: Cambridge University Press.
- Chaiken, S.** (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752–766. DOI: <https://doi.org/10.1037/0022-3514.39.5.752>
- Cohen, J.** (1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312. DOI: <https://doi.org/10.1037/0003-066X.45.12.1304>
- Cook, T. D., Campbell, D. T., & Peracchio, L.** (1990). Quasi Experimentation. In: Dunnette, M. D., & Hough, L. M. (Eds.), *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press.
- Deutsch, M., & Krauss, R. M.** (1965). *Theories in Social Psychology*. NY: Basic Books.
- Fiedler, K., Schott, M., & Meiser, T.** (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47*, 1231–1236. DOI: <https://doi.org/10.1016/j.jesp.2011.05.007>
- Finkel, E. J., Eastwick, P. W., & Reis, H. T.** (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology, 108*, 275–297. DOI: <https://doi.org/10.1037/pspi0000007>
- Flake, J. K., Pek, J., & Hehman, E.** (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*. DOI: <https://doi.org/10.1177/1948550617693063>
- Foschi, M.** (2014). Hypotheses, operationalizations, and manipulation checks. In: Webster, M. J., & Sell, J. (Eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. (pp. 247–268). San Diego, CA, US: Elsevier Academic Press. DOI: <https://doi.org/10.1016/B978-0-12-404681-8.00011-X>
- Haslam, S. A., & McGarty, C.** (2003). Experimental design and causality in social psychological research. In: Sansone, C., Morf, C. C., & Panter, A. T. (Eds.), *Handbook of Methods in Social Psychology*, (pp. 237–264), Thousand Oaks, CA: Sage.
- Hüffmeier, J., Mazei, J., & Schultze, T.** (2016). Reconceptualizing replication as a sequence of different

- studies: A replication typology. *Journal of Experimental Social Psychology*, 66, 81–92. DOI: <https://doi.org/10.1016/j.jesp.2015.09.009>
- Jacoby, J., & Sassenberg, K.** (2011). Interactions do not only tell us *when*, but can also tell us *how*: Testing process hypotheses by interaction. *European Journal of Social Psychology*, 41. DOI: <https://doi.org/10.1002/ejsp.762>
- John, L. K., Loewenstein, G., & Prelec, D.** (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532. DOI: <https://doi.org/10.1177/0956797611430953>
- Kühnen, U.** (2010). Manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin*, 36, 47–58. DOI: <https://doi.org/10.1177/0146167209346746>
- Maxwell, S. E.** (2004). The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies. *Psychological Methods*, 9, 147–163. DOI: <https://doi.org/10.1037/1082-989X.9.2.147>
- Messick, S.** (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14, 5–8. DOI: <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- O’Keefe, D. J.** (2003). Message Properties, Mediating States, and Manipulation Checks: Claims, Evidence, and Data Analysis in Experimental Persuasive Message Effects Research. *Communication Theory*, 13, 251–274. DOI: <https://doi.org/10.1111/j.1468-2885.2003.tb00292.x>
- Platt, J. R.** (1964). Strong inference. *Science*, 146, 347–353. DOI: <https://doi.org/10.1126/science.146.3642.347>
- Reis, H. T., & Judd, C. M.** (2014). *Handbook of Research Methods in Social and Personality Psychology*, 2nd ed. New York: Cambridge University Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T.** (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA, US: Houghton, Mifflin and Company.
- Sigall, H., & Mills, J.** (1998). Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personality and Social Psychology Review*, 2, 218–226. DOI: https://doi.org/10.1207/s15327957pspr0203_5
- Stroebe, W., & Strack, F.** (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9, 59–71. DOI: <https://doi.org/10.1177/1745691613514450>
- Wilson, T. D., Aronson, E., & Carlsmith, K.** (2010). The art of laboratory experimentation. In: Fiske, S. T., Gilbert, D. T., & Lindzey, G. (Eds.), *Handbook of Social Psychology*, 1, 51–81. NJ: Wiley. DOI: <https://doi.org/10.1002/9780470561119.socpsy001002>

How to cite this article: Fayant, M.-P., Sigall, H., Lemonnier, A., Retsin, E., & Alexopoulos, T. (2017). On the Limitations of Manipulation Checks: An Obstacle Toward Cumulative Science. *International Review of Social Psychology*, 30(1), 125–130, DOI: <https://doi.org/10.5334/irsp.102>

Published: 20 June 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

[u] *International Review of Social Psychology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 