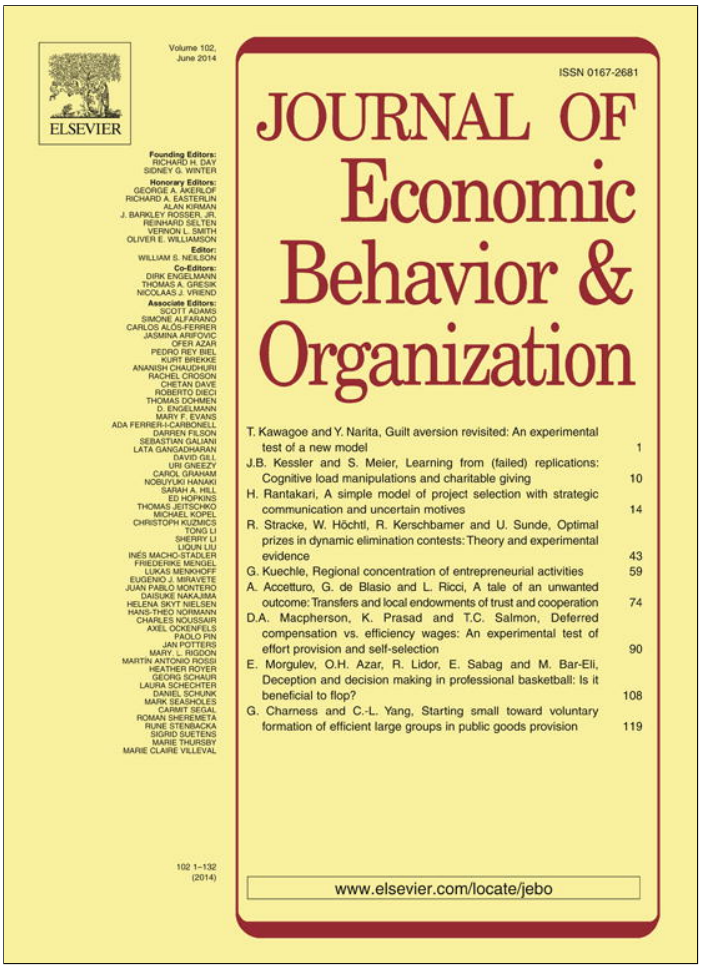

Journal of Economic Behavior & Organization 102 (2014) 10–13

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Learning from (failed) replications: Cognitive load manipulations and charitable giving☆

Judd B. Kessler[a], Stephan Meier[b,*]

[a] University of Pennsylvania, United States
[b] Columbia University, United States

ARTICLE INFO

Article history:
Received 15 December 2013
Accepted 6 February 2014
Available online 5 March 2014

Keywords:
Methodology
Cognitive load
Charitable giving

ABSTRACT

Replication of empirical studies is much more than a tool to police the field. Failed replications force us to recognize that seemingly arbitrary design features may impact results in important ways. We describe a study that used a cognitive load manipulation to investigate the role of the deliberative system in charitable giving and a set of failed replications of that study. While the original study showed large and statistically significant results, we failed to replicate using the same protocol and the same subject pool. After the first failed replication, we hypothesized that the order our study was taken in a set of unrelated studies in a laboratory session generated the differences in effects. Three more replication attempts supported this hypothesis. The study demonstrates the importance of replication in advancing our understanding of the mechanisms driving a particular result and it questions the robustness of results established by cognitive load tests.

1. Introduction

Replication of empirical studies can help identify false positive results and uncover questionable research practices (see special section on replicability in *Perspectives in Psychological Science*, including Pashler and Wagenmakers, 2012). But running replications is about more than policing the field. Replications can elucidate how subtle differences in setting, subject pool, and protocol impact results. Specifically, failed replications force us to recognize that some seemingly arbitrary design features may be necessary for a result to arise, which can help us understand the mechanisms driving the effect (for a similar argument about laboratory and field experiments see Kessler, 2013).

This paper describes a set of studies in which we use a cognitive load manipulation to investigate the role of the deliberative system in charitable giving (for a discussion of mental processes on charitable giving, see Loewenstein and Small, 2007). Our original results suggested that people gave substantially more money to a charity when placed under high cognitive load, results that were consistent with other findings in the literature. Schulz et al. (in press) and Rand et al. (2012) have found similar effects; however, Hauge et al. (2009) found no impact of load on giving.

Given our large treatment effect and statistically significant results, we were confident that our findings had pushed forward the frontiers of knowledge (see manuscript in SOM). While we were conducting more research, however, we failed

☆ We thank Theresa Kelly for excellent work. We thank The Wharton School for research support and the Wharton Behavioral Lab for administering the studies.

* Corresponding author at: Columbia University, Graduate School of Business, 710 Uris Hall, 3022 Broadway, New York, NY 10027, United States. Tel.: +1 212 851 1801.
E-mail address: sm3087@columbia.edu (S. Meier).

http://dx.doi.org/10.1016/j.jebo.2014.02.005
0167-2681/

to replicate our original result. We hypothesized that a subtle difference between the original study and the failed replication attempt—the order our study was taken in an hour-long laboratory session—generated the difference in results. Three more replication attempts supported this hypothesis.

2. Method

The original study was a 2 (charity request or general request) × 2 (low or high cognitive load) between-subjects design. The replication attempts kept the same design, but here we focus only on the charity requests (see SOM for results from all sessions). Each study was one of several unrelated studies in an hour-long session at the Wharton Behavioral Lab. Sometimes our study was first in the session and sometimes it followed other studies. We were initially indifferent about the session order. Subsequently, we explicitly asked to be at the start of the session or fourth in the session as explained below.

2.1. Subjects

We analyze 405 University of Pennsylvania undergraduates (53.3% female) who participated in our charity request condition. Subjects received $10 payment for the hour-long session and whatever amount they chose to keep in our study. Across all studies, the subject pool and the instructions were kept the same.

2.2. Charity request

Subjects were placed under high or low cognitive load, given an endowment of $3, and asked how much of their endowment they wanted to donate to the American Red Cross.

2.3. Cognitive load manipulation

Cognitive load manipulations often involve memorizing a sequence (e.g., Gilbert et al., 1995; Gilbert and Osborne, 1989; Shiv and Fedorikhin, 1999). Since subjects made numeric giving decisions, we used a sequence of letters to avoid anchoring effects (Tversky, 1974). Subjects were randomly assigned to memorize a 3-letter sequence ("GXN") (low cognitive load), or a 9-letter sequence ("GXNTDPLRW") (high cognitive load). We did not incentivize the load manipulation to avoid income effects.

3. Results

Fig. 1 shows donations by cognitive load condition in the original study and in four replication attempts. In the original study subjects under high load give twice as much ($0.51 vs. $1.12; 102 obs; t-test, $t = 2.99$, $p = 0.004$) and are 50% more likely to give (38% vs. 58%; 102 obs; pr-test, $z = 1.97$, $p = 0.048$). In Replications 1, 2 and 4 the sign of the effect is reversed and the effect on average donation is statistically significantly different from the effect in the original study ($p < 0.05$ for all tests). We only replicate in attempt 3: under high load probability of donation increases (60% vs. 79%; 101 obs; pr-test $z = 2.00$, $p = 0.046$) and average donation directionally increases ($1.03 vs. $1.35, t-test, $t = 1.39$, $p = 0.168$).

4. Order in session

After our first failed replication attempt, we hypothesized that the difference in results was due to when our study was conducted in the hour-long lab session. The original study was run fourth and Replication 1 was run first. Starting with Replication 2, we specifically asked to be run either first (Replication 2 and Replication 4) or fourth (Replication 3).

Our data confirms our hypothesis that session order matters. When our study is first in the session, cognitive load directionally *reduces* charitable giving ($0.97 vs. $0.78); when it is later in the session, cognitive load *increases* charitable giving ($0.78 vs. $1.23; 203 obs; t-test, $t = 2.84$, $p = 0.005$). The effect of load on giving statistically significantly interacts with session order (405 obs; OLS on average donation $p = 0.004$; OLS on probability of donation $p = 0.020$). The effect gets stronger when we control for the calendar date on which a session was run, allowing subjects who participate on different dates to have different baseline levels of generosity (405 obs; OLS on average donation $p = 0.001$; OLS on probability of donation $p = 0.006$).

This pattern of results holds when looking only at the three replications conducted after forming our session-order hypothesis (233 obs; without date controls: OLS on average donation $p = 0.053$; OLS on probability of donation $p = 0.024$; with date controls: OLS on average donation $p = 0.030$; OLS on probability of donation $p = 0.008$).

5. Discussion

What can we learn by comparing an original study to its failed replications? When another researcher fails to replicate a study, the lack of a result might arise from differences in methods, subject pool, environment, or some other factor (see, e.g., Harless, 1992, for a study where a subtle difference in the presentation of a choice problem significantly changes the

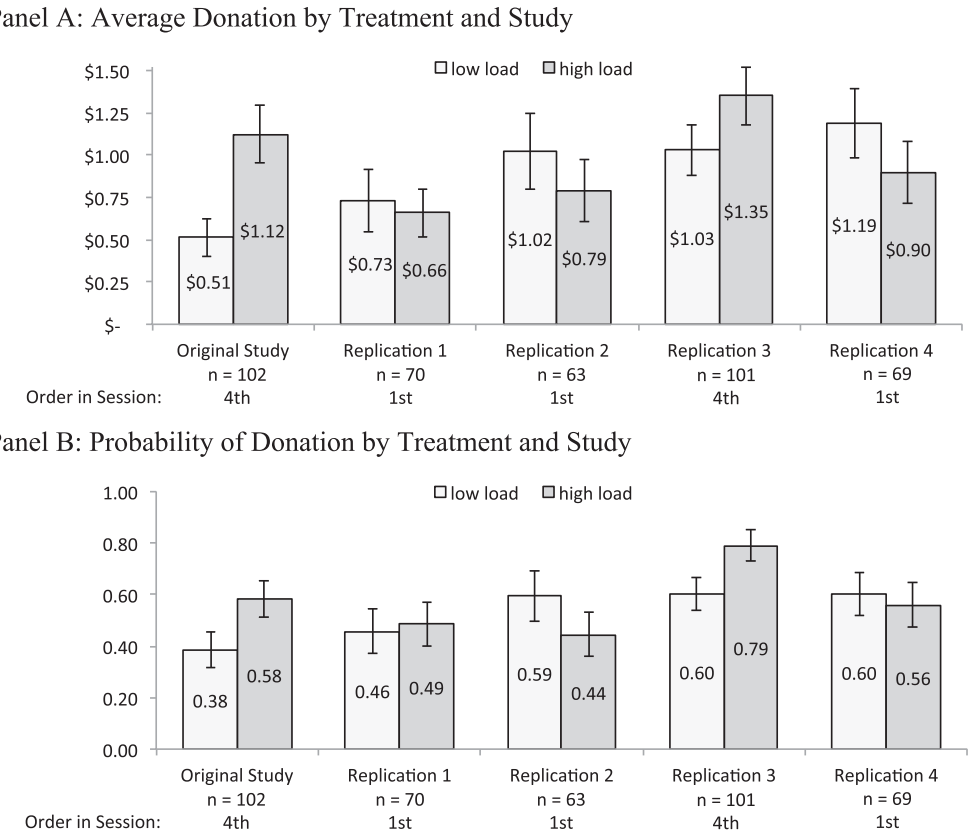

Fig. 1. Original study and four replication attempts. Panel A shows average amount of the $3 endowment donated to the American Red Cross (means ± SEM). Panel B shows percent of subject who donated a positive amount to the American Red Cross (±SEM).

results). When a researcher fails to replicate a result using the same instructions, subject pool, and laboratory environment, one must look for subtle differences between the replication attempt and the original study.

Whether our study is first or later in a session affects the *sign* of the effect of cognitive load on charitable giving. One possible explanation is that the efficacy of our cognitive load manipulation is sensitive to session order. For example, the cognitive load task might more completely occupy the deliberative system if subjects have suffered mental fatigue from participating in studies earlier in the session. Similarly, subjects might be differently inclined to spend mental energy remembering the long string of letters in the high load treatment depending on mental fatigue. Our manipulation check questions provide some evidence that cognitive load is differently effective early and late in a session (see SOM4). Another possible explanation is that load is equally effective early and late in a session but that its effect on giving may be heavily moderated by the context of the request—even context as subtle as when in a session of studies the subject is asked to donate.

Both of these possible explanations are worthy of future study. The former explanation tells a very cautious tale about cognitive load manipulations and suggests further replication attempts of results established using cognitive load manipulations. Both explanations suggest that role of the deliberative system in charitable giving is far from resolved and that one should be cautious in relying on cognitive load manipulations to establish that result. A similar debate has arisen about the effectiveness of time–pressure tests on the role of the deliberative system in charitable giving (Rand et al., 2012; Tinghög et al., 2013).

A natural next step for future research would be to randomize when in a session a load manipulation is run (rather than relying on between-session data) to investigate whether the treatment effect changes monotonically with session order and to investigate whether the types of other tasks subjects complete in a session moderate the effect of load on giving.

Overall, these results demonstrate an important value of replication. Even when a replication fails, it may be able to teach something about the original effect.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2014.02.005.

References

Gilbert, D.T., Giesler, R.B., Morris, K.A., 1995. When comparisons arise? Journal of Personality and Social Psychology 69 (2), 227–236.

Gilbert, D.T., Osborne, R.E., 1989. Thinking backward: some curable and incurable consequences of cognitive busyness. Journal of Personality and Social Psychology 57 (6), 940–949.
Harless, D.W., 1992. Actions versus prospects: the effect of problem representation on regret. American Economic Review 82 (3), 634–649.
Hauge, K.E., Brekke, K.A., Johansson, L., Johansson-Stenman, O., Svedsäter, H., 2009. Are Social Preferences Skin Deep? Dictators Under Cognitive Load (Working Paper No. 371). Retrieved from the Gothenburg University Publications Electronic Archive: http://hdl.handle.net/2077/20768
Kessler, J.B., 2013. When Will There be Gift Exchange? Addressing the Lab-Field Debate with Laboratory Gift Exchange Experiments (CESifo Working Paper No. 4161).
Loewenstein, G., Small, D., 2007. The scarecrow and the tin man: the vicissitudes of human sympathy and caring. Review of General Psychology 2 (1), 112–126.
Pashler, H., Wagenmakers, E. (Eds.), 2012. Special section on replicability in psychological science: a crisis of confidence? Perspectives on Psychological Science 7 (6), 528–654 (Special section).
Rand, D.G., Greene, J.D., Nowak, M.A., 2012. Spontaneous giving and calculated greed. Nature 489 (7416), 427–430.
Schulz, J., Fischbacher, U., Thoeni, C., Utikal, V., 2014. Affect and fairness: dictator games under cognitive load. Journal of Economic Psychology (in press).
Shiv, B., Fedorikhin, A., 1999. Heart and mind in conflict: the interplay of affect and cognition in consumer decision making. Journal of Consumer Research 26 (3), 278–292.
Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., Västfjäll, D., Kirchler, M., Johannesson, M., 2013. Intuition and cooperation reconsidered. Nature 498, E1–E2.
Tversky, A., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.