Journal of Experimental Social Psychology 69 (2017) 144–149

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

The “true” indirect effect won't (always) stand up: When and why reverse mediation testing fails

Gunnar Lemmer *, Mario Gollwitzer

Department of Psychology, Philipps-University Marburg, Germany

HIGHLIGHTS

- Tests of mediation hypotheses (i.e., $X \rightarrow M \rightarrow Y$) are widespread in social psychology
- In many cases, "$X \rightarrow Y \rightarrow M$" is a plausible alternative model
- A common practice is to compare both indirect effects (reverse mediation testing)
- Monte Carlo simulations show that this approach often leads to wrong conclusions
- This especially applies when "M" is measured less reliably than "Y"

ARTICLE INFO

Article history:
Received 23 July 2015
Revised 5 February 2016
Accepted 10 May 2016
Available online 1 June 2016

Keywords:
Mediation processes
Causality
Reverse mediation
Monte Carlo simulation

ABSTRACT

Many social psychological studies aim to test whether an independent variable (X) affects a dependent variable (Y) via one (or more) intervening variable(s) or "mediator(s)" (M). One way to test such a mediation model ($X \rightarrow M \rightarrow Y$) is to manipulate X, measure both M and Y, and test statistically whether the indirect effect of X on Y via M is significantly different from zero. However, since the causal order between M and Y is unclear, alternative models (such as $X \rightarrow Y \rightarrow M$) are also compatible with the data. Scholars have argued that comparing such models statistically against each other can help decide which model is “correct.” In the present article, we scrutinize the tenability of this “reverse mediation testing” approach via Monte Carlo simulations. Our findings show that reverse mediation testing often fails—especially when the mediator is measured less reliably than the dependent variable.

—“*... the mediational model, in which variable M is presumed to mediate the effect of X (the treatment) on O (the outcome). . . . Instead of this model, one might estimate the second model, switching the roles of M and O. Hopefully, this would lead to the conclusion that O fails to mediate the X – M relationship.*”

(Judd & Sadler, 2008, pp. 129–130)

—“*Often it is advisable to interchange the mediator and the outcome variable and have the outcome* “cause” *the mediator. If the results look similar to the specified mediational pattern. . . , one would be less confident in the specified model. However, it should be realized that the direction of causation between M and Y cannot be determined by statistical analyses.*”

(Kenny, n.d.; retrieved from http://davidakenny.net/cm/mediate.htm).

Many social psychological theories assume that an independent variable (e.g., an intervention) influences a dependent variable (e.g., the assumed outcome) through its effect on one (or multiple) intervening variable(s). In other words, the independent variable (X) is assumed to have an indirect effect on the dependent variable (Y) via one (or more) "mediator(s)" (M). Testing mediation hypotheses can help illuminating the psychological mechanism underlying a causal effect of X on Y (Bullock, Green, & Ha, 2010). For instance, a researcher may assume that treating participants unfairly (vs. fairly) makes them act vengefully against the perpetrator because the unfair treatment has increased participants' moral outrage. This hypothesis implies that moral outrage mediates the effect of unfair (vs. fair) treatment on vengeful behavior.

* Corresponding author at: Department of Psychology, Section for Psychological Methods and Social Psychology, Philipps-University Marburg, Gutenbergstrasse 18, 35032 Marburg, Germany.
E-mail address: gunnar.lemmer@uni-marburg.de (G. Lemmer).

http://dx.doi.org/10.1016/j.jesp.2016.05.002

The social psychological literature is full of mediation hypotheses (and tests of them). A closer inspection of the full texts of papers published in three of the top-tier social psychological journals reveals that almost half of all empirical papers published in 2014 in the *Journal of Experimental Social Psychology* (JESP; 44%), the *Personality and Social Psychology Bulletin* (PSPB; 44%), and the *Journal of Personality and Social Psychology* (JPSP; 47%) include at least one test of a mediation hypothesis.

Mediation hypotheses can be tested with different research designs. The assumed mediator can be manipulated ("experimental causal chaining;" see Spencer, Zanna, & Fong, 2005, or "mediation by moderation;" see Jacoby & Sassenberg, 2011) or measured just as the dependent variable. Most often, the latter approach is used. Then, traditionally (see Baron & Kenny, 1986), "the total effect" of X on Y is compared against the "direct effect" (i.e., the effect of X on Y after controlling for M; $b_{12} \times b_{23}$ in Fig. 1). For instance, after randomly assigning participants to either unfair or fair treatment, the experimenter could measure participants' moral outrage (M) as well as their vengeful behavior (Y) and test whether the effect of unfair vs. fair treatment (X) on revenge is significantly reduced after statistically controlling for self-reported moral outrage. This so-called "causal steps" approach (Baron & Kenny, 1986) has been discussed extensively in the literature (e.g., Hayes, 2013; Jose, 2013; MacKinnon, 2008; Rucker, Preacher, Tormala, & Petty, 2011). A related approach to probe a mediation hypothesis is to test the indirect effect of X via M on Y (i.e., $b_{12} \times b_{23}$ in Fig. 1) against zero. Since Sobel's (1982) introduction of a test statistic that assumes the sampling distribution of indirect effects to be normal (which is rarely the case), novel methods to test indirect effects for statistical significance have been introduced and are now widely applied (e.g., Fritz, Taylor, & MacKinnon, 2012; Hayes & Scharkow, 2013).

A major problem is that a statistically significant indirect effect does not allow to rule out alternative models (Danner, Hagemann, & Fiedler, 2015; Fiedler, Schott, & Meiser, 2011; MacKinnon & Pirlott, 2015). For instance, the nominal mediator might be just a correlate of the "true" (but unmeasured) mediator, or the $M \rightarrow Y$ path (i.e., b_{23}) might just represent a spurious correlation that is caused by a hidden variable. These and further alternative models are equally plausible and cannot easily be tested against each other (but see Danner et al., 2015).

In the present article, we will focus on one specific alternative model "$X \rightarrow Y \rightarrow M$". This alternative model assumes that the independent variable X (e.g., treating participants unfairly vs. fairly) *directly* influences the dependent variable Y (e.g., vengeful behavior), and that this variable, in turn, causally influences the nominal mediator variable M (e.g., self-reported moral outrage about the unfairness). Psychologically, such a model may be as plausible as the originally hypothesized model. Notably, with a statistically significant indirect effect "$X \rightarrow M \rightarrow Y$"—irrespective of the concrete statistical method—researchers cannot differentiate between the two models per se. One way to deal with this problem has been introduced as the "reverse mediation testing" (RMT) approach (see the two quotes from the beginning of this article). In this approach, the indirect effect in the "target model" ($X \rightarrow M \rightarrow Y$) is compared against the respective indirect effect in the "alternative model" ($X \rightarrow Y \rightarrow M$). The

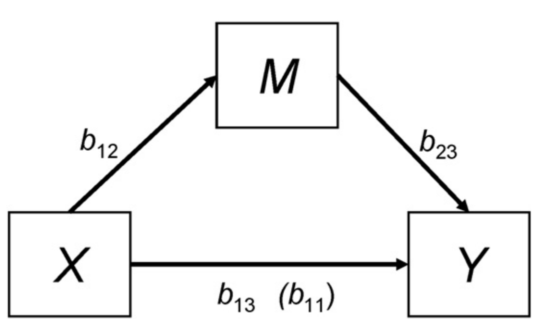

Fig. 1. Mediation model with three variables.

rationale underlying this approach is the following: if (and only if) the target model is correct, then the indirect effect $b_{12} \times b_{23}$ should be considerably larger than the indirect effect in the alternative model. A usual practice is to favor the model in which the indirect effect is statistically significant and to reject the model in which the indirect effect is not significant.

The RMT approach is intuitively plausible and easy to implement. But is it really useful to improve our conclusions about causal processes or can it even lead to erroneous inferences? The aim of the present research is to elucidate whether RMT really leads to a valid conclusion, and we will argue that it often does not. We are not the first to argue that RMT is flawed. Thoemmes (2015) has recently used an analytic approach to show that comparing the size of indirect effects in alternative models often leads to erroneous decisions. He demonstrates that given a true indirect effect, it is quite possible that alternative models have larger indirect effects. Thoemmes (2015) concludes that "Reversing arrows in mediation models does not tell us whether one model is better than the other" (p. 230). In the present paper, we go one step further by also considering statistical significance and exploring the conditions under which RMT fails with Monte Carlo simulations. We will show that measurement error, that is, the unreliability of measured variables Y and M, plays a major role for the fact that RMT often fails.

1. Measurement error

In our aforementioned example, the independent variable (i.e., unfair vs. fair treatment) is varied under full experimental control; so X does not contain any measurement error. However, both the assumed mediator M (moral outrage) as well as the dependent variable Y (vengeful behavior) are probably measured with error. Remember that the indirect effect of X on Y is the product of two direct effects: the direct effect of X on M (i.e., b_{12} in Fig. 1), and the direct effect of M on Y (i.e., b_{23}), which is estimated in a multiple regression model where Y is regressed simultaneously on both M and X. Consequently, the extent to which M contains measurement error can bias the results (Baron & Kenny, 1986; Hoyle & Robinson, 2004). Conversely, in the alternative model, the indirect effect is calculated as the product of the direct effect of X on Y and the direct effect of Y on M. Accordingly, the unreliability of Y can lead to biased results. Hence, different findings for the indirect effects in the target and the alternative model might be merely the result of measurement error (i.e., the unreliability of M and Y). In the present paper, we will demonstrate that measurement error can seriously bias the conclusions one can draw from the "reverse mediation testing" approach.

One solution to this problem could be to test the indirect effect $X \rightarrow M \rightarrow Y$ with *latent* variables. Latent variable (or "structural equation") modeling allows researchers to separate true-score from error variance; thus, testing an indirect effect in a structural equation model has the advantage of avoiding measurement error. In fact, many scholars have argued that a mediation hypothesis should preferably be tested with a structural equation model (e.g., Bollen, 1989; Brown, 1997; Cheung & Lau, 2008). These authors correctly argue that structural equation models reduce the problem of artificial biases in mediation analyses (Hoyle & Smith, 1994). But does structural equation modeling also guarantee valid conclusions about causal effects if the "reverse mediation testing" (RMT) approach is used? This will be explored in the present paper.

2. Overview of the present analysis

The research presented here critically scrutinizes the RMT approach—both on the manifest and on the latent level. More precisely, we will describe the results of a Monte Carlo simulation study. In a Monte Carlo simulation (see Paxton, Curran, Bollen, Kirby, & Chen, 2001) a large number of random samples of a given size are drawn from a population for which a specified model applies. Then, the sample

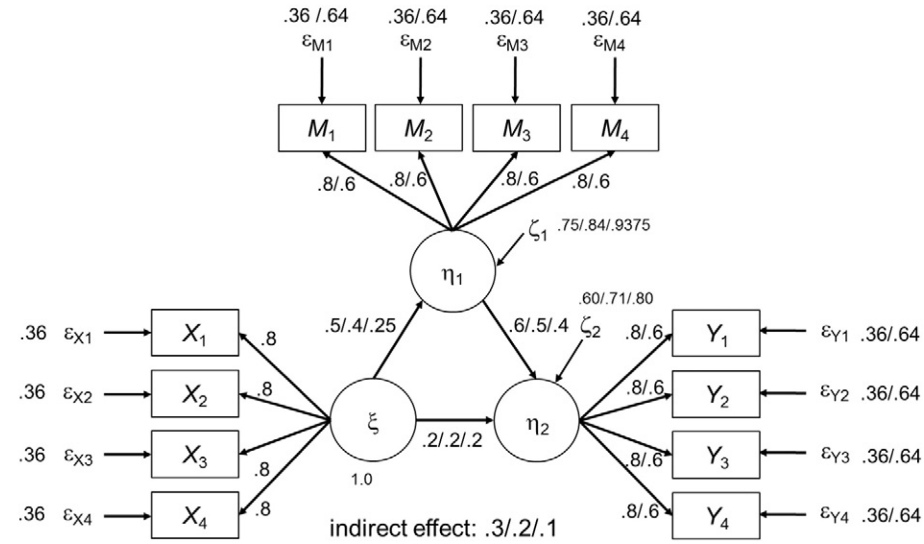

Fig. 2. Illustration of the Population Models. ξ = latent exogenous variable representing X; η_1 = latent mediator variable representing M; η_2 = latent endogenous variable representing Y; ζ_1 = latent residual variable for η_1; ζ_2 = latent residual variable for η_2; ε_i = residual variable for manifest variable i. Coefficients separated by a slash were varied in different population models.

data are analyzed with regard to specific aspects. In the current work, we conducted multiple simulations. In each simulation, we specified a population for which a specific assumed model applies, drew a large number of samples from each population, fitted both the target and the alternative mediation model to the sample data, and compared the indirect effect in both models. We were interested in the extent to which this comparison depends on (1) the (un)reliability of *M* and *Y* in the population model, (2) the size of the indirect effect in the population model, (3) the sample size, and (4) the type of analysis (i.e. manifest or latent).

3. Method

In order to reduce complexity, we refer to a simple "tri-variate" mediation model with one independent variable, one mediator variable, and one dependent variable. In our simulation study, each construct is measured with four indicators.[1] The analysis was conducted in four steps.

In the first step, we specified *population models* with latent variables (see Fig. 2). In these structural equation models, the latent independent variable ξ (indicated by four manifest variables X_1–X_4) has an indirect effect on the latent dependent variable η_2 (indicated by Y_1–Y_4) via the latent mediator η_1 (indicated by M_1–M_4). In addition, a direct effect of ξ on η_2 was specified. Several subtypes of this basic population model were formulated to vary the reliability of the indicator variables for η_1 and η_2 as well as the size of the indirect effect in the population. All variables were specified to have a mean of 0 and a standard deviation of 1 in the population. As one aim of our study is to test the influence of measurement error, the reliability of the indicators was varied by fixing their loadings on the respective factor to 0.8 or 0.6, so that either 20% or 40% of their observed variance is due to measurement error (and indicator-specific true variance). Altogether, we specified three configurations: (1) all indicator variables have a factor loading of 0.8 and are equally reliable; (2) the indicators of the dependent variable are less reliable with factor loadings of 0.6; and (3) the indicators of the mediator are less reliable with factor loadings of 0.6. Also, the standardized indirect effect was modeled to be 0.1, 0.2, or 0.3 (see Fig. 1). The choice of these specific values was arbitrary, but not implausible for social psychological research. To sum up, 3 (reliability of indicator variables) × 3 (size of indirect effect) = 9 specific population models were specified.

The sample data were simulated in the second step. From each of the nine population scenarios, we drew 1000 samples with $N = 50$, $N = 100$, or $N = 200$ elements and normally distributed indicator variables for *X*, *M*, and *Y*. Sample size was varied to investigate the influence of statistical power in a broader way and not only with regard to the size of the indirect effect in the population. Altogether, 9 (specific population model) × 1000 (samples) × 3 (sample size) = 27,000 data sets were generated.

In the third step, we fitted four *sample models* to each of the generated data sets: manifest target, reverse manifest, latent target, and reverse latent. For the *manifest target model*, *X*, *M*, and *Y* were calculated as arithmetic means of their indicators. In this path model, the causal order is expressed with manifest variables: *X* has an indirect effect on *Y* via *M* and a direct effect on *Y*. In the *reverse manifest model*, the location of *M* and *Y* is switched. With this model, it was wrongly assumed that *X* had an indirect effect on *M* through *Y* and a direct effect on *M*. We also specified a *latent target model* in which *X*, *M*, and *Y* were modeled as latent variables ξ, η_1, and η_2. It resembles the population model, but to mimic prototypical data analyses conducted in social psychological research, the factor loadings and path coefficients were freed from constraints; only the necessary scaling restrictions were made. Finally, in the *reverse latent model*, *X*, *M*, and *Y* were also specified as latent variables ξ, η_1, and, η_2; however, it was erroneously assumed that *Y* was the mediator and *M* the outcome variable.

In the fourth and final step, we aggregated the sample findings. More specifically, across the 1000 sample results obtained for each of the 9 (specific population model) × 3 (sample size) × 4 (fitted sample model) = 108 configurations, we determined the average of the estimated indirect effect, the standard deviation of the estimates, the average of the estimated standard errors, and the percentage of cases with a statistically significant indirect effect[2] ($\alpha = 0.05$). The percentage of significant indirect effects can be interpreted as the probability of obtaining a significant indirect effect given the specific configuration. All analyses were conducted with M*plus* 7.0 (Muthén & Muthén, 1998–2012).

4. Results

The results of our simulation study are displayed in Tables 1–3. Each table displays average standardized indirect effects (SIE; aggregated

[1] An experimental scenario in which *X* is manipulated and *M* and *Y* is measured is a special case of the general framework we refer to in our analyses. In such a special case, *X* can be seen as a single-indicator of the DV that is free of measurement error.

[2] These findings are based on the Sobel Test since the bootstrapping approach is not implemented in the Monte Carlo simulation facility of M*plus*. However, differences between both approaches can be expected to be negligible.

Table 1
Findings obtained when the standardized indirect effect in the population is specified to be 0.3.

		Sample size of the individual data sets					
		$N = 200$		$N = 100$		$N = 50$	
Reliability	Sample model	Average SIE	% sig.	Average SIE	% sig.	Average SIE	% sig.
$M = Y$	Manifest	0.216	1.000	0.214	0.995	0.213	0.824
	Manifest reverse	0.201	1.000	0.199	0.990	0.199	0.791
	Latent	0.282	0.999	0.280	0.994	0.280	0.802
	Latent reverse	0.257	0.999	0.255	0.991	0.255	0.759
$M > Y$	Manifest	0.194	1.000	0.192	0.991	0.191	0.735
	Manifest reverse	0.156	1.000	0.154	0.967	0.154	0.602
	Latent	0.282	0.999	0.279	0.982	0.280	0.687
	Latent reverse	0.258	0.999	0.258	0.952	0.264	0.507
$M < Y$	Manifest	0.162	1.000	0.160	0.967	0.160	0.631
	Manifest reverse	0.179	1.000	0.177	0.983	0.177	0.690
	Latent	0.283	0.999	0.285	0.942	0.297	0.481
	Latent reverse	0.257	0.999	0.255	0.975	0.255	0.641

Note. Average SIE = average standardized indirect effect across 1000 samples; % sig. = percentage of significant indirect effects ($\alpha = 0.05$).

across 1000 samples each) together with the proportion of samples in which the respective effect was significant (based on $\alpha = 0.05$) as a function of (a) size of the standardized indirect effect in the population (in different tables), (b) sample size (in columns), (c) reliability of indicators in the population (in blocks of rows), and (d) the model fitted to the sample data (in rows), respectively. A more detailed presentation of the results can be found in the Supplementary Online Material.

Let us illustrate our findings with a case that is presumably typical for many social psychological applications: the indirect effect in the population model is 0.2 and the sample size is $N = 100$ (see Table 2). Within each of the three relative reliability categories (represented as blocks of rows), the first row shows the results for analyses in which the sample model was correctly specified (i.e., $X \rightarrow M \rightarrow Y$) and effects were tested on the level of observed variables (*manifest target model*). Here, we would expect a large number of significant indirect effects. The second row shows the results for a manifest sample model in which the reverse mediation (i.e., $X \rightarrow Y \rightarrow M$) was specified (*manifest reverse model*). If the RMT approach is useful, then we would expect (a) a smaller average point estimate of the indirect effect and (b) a smaller number of significant indirect effects than for the respective target model. The third and fourth rows refer again to the target and the reverse mediation model, but this time the effects were tested on the level of latent variables (*latent target model* and *latent reverse model*).

Table 2
Findings obtained when the standardized indirect effect in the population is specified to be 0.2.

		Sample size of the individual data sets					
		$N = 200$		$N = 100$		$N = 50$	
Reliability	Sample model	Average SIE	% sig.	Average SIE	% sig.	Average SIE	% sig.
$M = Y$	Manifest	0.148	1.000	0.146	0.914	0.145	0.527
	Manifest reverse	0.142	1.000	0.140	0.898	0.139	0.501
	Latent	0.191	0.999	0.189	0.912	0.189	0.498
	Latent reverse	0.182	0.999	0.179	0.901	0.178	0.477
$M > Y$	Manifest	0.132	1.000	0.130	0.873	0.193	0.691
	Manifest reverse	0.111	0.996	0.109	0.779	0.129	0.496
	Latent	0.191	0.999	0.189	0.858	0.274	0.626
	Latent reverse	0.182	0.995	0.181	0.759	0.209	0.402
$M < Y$	Manifest	0.113	0.997	0.111	0.770	0.111	0.328
	Manifest reverse	0.126	1.000	0.124	0.852	0.124	0.405
	Latent	0.192	0.996	0.191	0.741	0.196	0.252
	Latent reverse	0.181	0.999	0.179	0.841	0.178	0.387

Note. Average SIE = average standardized indirect effect across 1000 samples; % sig. = percentage of significant indirect effects ($\alpha = 0.05$).

Table 3
Findings obtained when the standardized indirect effect in the population is specified to be 0.1.

		Sample size of the individual data sets					
		$N = 200$		$N = 100$		$N = 50$	
Reliability	Sample model	Average SIE	% sig.	Average SIE	% sig.	Average SIE	% sig.
$M = Y$	Manifest	0.076	0.849	0.074	0.448	0.074	0.146
	Manifest reverse	0.091	0.952	0.089	0.581	0.089	0.215
	Latent	0.098	0.857	0.095	0.444	0.095	0.142
	Latent reverse	0.118	0.953	0.115	0.586	0.115	0.195
$M > Y$	Manifest	0.068	0.826	0.066	0.363	0.066	0.118
	Manifest reverse	0.072	0.867	0.070	0.409	0.069	0.126
	Latent	0.097	0.833	0.095	0.363	0.095	0.114
	Latent reverse	0.118	0.863	0.117	0.387	0.117	0.082
$M < Y$	Manifest	0.059	0.707	0.057	0.297	0.057	0.083
	Manifest reverse	0.081	0.934	0.079	0.503	0.079	0.157
	Latent	0.098	0.698	0.096	0.269	0.097	0.072
	Latent reverse	0.118	0.933	0.116	0.501	0.114	0.145

Note. Average SIE = average standardized indirect effect across 1000 samples; % sig. = percentage of significant indirect effects ($\alpha = 0.05$).

When indicators of M and Y are equally reliable, the average point estimate of the indirect effect is only slightly larger in the manifest and latent target models (0.146 and 0.189, respectively) than in the corresponding reverse manifest and latent models (0.14 and 0.179, respectively). In addition, the average point estimate is larger on the latent than on the manifest level. With regard to statistical significance, the null hypothesis stating that the indirect effect is zero in the population could be rejected in 90% to 91% of the samples irrespective of the sample model. Hence, the RMT approach does not yield clear results; the chance of finding a significant reverse indirect effect is virtually identical to the chance of obtaining a significant indirect effect in the target model.

When indicators of Y are measured less reliably than indicators of M, however, the chance of finding a significant reverse indirect effect is indeed lower (i.e., 78% and 76% for manifest and latent variable modeling, respectively) than the chance of finding a significant target indirect effect (87% and 86%, respectively). Also, reverse indirect effects are smaller in size than target indirect effects, at least for manifest variable modeling. Here, the RMT approach seems to yield useful results.

Finally, when indicators of M are measured less reliably than indicators of Y, the chance of finding a significant reverse indirect effect is even higher (i.e., 85% and 84% for manifest and latent variable modeling, respectively) than the chance of finding a significant target indirect effect (77% and 74%, respectively). In addition, reverse indirect effects are larger in size than target indirect effects, at least on the manifest level. Thus, when indicators of M are measured more unreliably than indicators of Y, the RMT approach is problematic: models testing the reverse mediation are more likely to (a) yield larger indirect effects and (b) become significant than models testing the target mediation effect. This might lead researchers to prefer the "wrong" model over the "right" one. Wow!

In all cases, the findings for latent models resemble those for manifest models. Interestingly, the results for latent indirect effects also depend on the reliability of the indicators. Also, the chance of finding a significant indirect effect is similar in the latent and manifest models.

The described pattern is generally replicated in the other scenarios. However, there is one exception: When the indirect effect is 0.1 in the population (see Table 3), when Y and M are equally reliable or the indicators of Y are less reliable, and when $N = 200$, $N = 100$, or $N = 50$, the average indirect effect and the percentage of significant indirect effects are higher for the reverse model than for the target model. This again applies to both the manifest and latent analysis and demonstrates that RMT can be misleading.

5. Discussion

In this article, we scrutinized the RMT approach according to which researchers can verify the tenability of their preferred mediation model (i.e., $X \rightarrow M \rightarrow Y$) by testing the size and the significance of the indirect effect against the indirect effect in a reverse mediation model (i.e., $X \rightarrow Y \rightarrow M$). More precisely, we presented the results of a Monte Carlo study in which we specified population models, sampled data sets from these populations, and fitted correct and reverse mediation models to the sample data. Our findings show that the RMT approach is generally not useful; only when *M* (the assumed mediator) is measured more reliably than *Y* (the dependent variable) it might be a justifiable strategy. When *Y* and *M* are equally reliable, the RMT approach does not really help distinguishing between the two alternative causal models: in both the target and the reverse mediation models, (a) the size of the indirect effects are similar, and (b) the indirect effects are equally likely to become significant.

The most important and notable conclusion that we can draw from our results is that the RMT fails when the mediator is less reliably measured than the DV. Here, the chance of finding a significant indirect effect in the reverse model is higher than in the target model. In addition, indirect effects in the reverse model are larger in size compared to target models. Hence, the RMT approach is not only dispensable in that case; it can even lead to inadequate conclusions.

As described in the results section, the tenability of the RMT approach is not only doubtful for manifest, but also for latent analyses. This finding is surprising: although the latent variables are themselves free of measurement error, the comparison between the indirect effect in the target and the reverse model depends on the reliability of the indicator variables, just as in manifest path models.

5.1. Should we refrain from statistical tests of mediation hypotheses?

Our research supports Thoemmes' (2015) statement that reversing arrows is not a useful strategy to test the tenability of a mediation model. We were able to demonstrate with simulated data that measurement error in *Y* and *M* plays a major role for the fact that the RMT approach is flawed and should be abandoned. But a possible reverse causal order is only one fragment of a bigger issue: many alternative models are viable explanations for a given data pattern (Fiedler et al., 2011). Given the fact that researchers can typically not rule out alternative explanations one may generally question the widespread practice of conducting mediation tests. As long as their limitations are not ignored and carefully considered in the interpretation of findings, we would not go that far. A recent paper by Danner et al. (2015) demonstrates that at least some alternative models can be ruled out statistically by comparing model fits. This strategy, however, can be only used for models that are overidentified in their structural part (i.e., at least one degree of freedom). The models we are most often interested in (i.e., tri-variate mediation models), however, are always just identified in their structural part. Overidentification in tri-variate mediation models only pertains the measurement model. Thus, inspection of model fit does not provide any information about the tenability of assumptions in the structural part. Although statistical modeling and the inspection of non-longitudinal data cannot clarify the direction of a mediation, it can improve mediation analysis by ruling out confounders (MacKinnon & Pirlott, 2015; see also other work on the causal inference framework: Imai, Tingley, & Yamamoto, 2010; Pearl, 2011, 2014; Valeri & VanderWeele, 2013).

Although we should generally not refrain from mediation analyses, they are often not justified. Before rushing to inspect statistical information about the "*empirical* tenability" of a mediation model, researchers should take one step back and carefully scrutinize the *conceptual* tenability of their model. Mediation hypotheses should be only specified (and tests of them should be only conducted) when the process "$X \rightarrow M \rightarrow Y$" is conceptually plausible. One important criterion to evaluate the conceptual tenability of a mediation model is the "conceptual timing criterion" (Tate, 2015; see also Hyman, 1955), which states that the values of *X* have to exist prior to the values of *M*, and the values of *M* have to exist prior to the values of *Y*. In other words, it has to be logically plausible that *X* changes *M* and that *M*, in turn, changes *Y*. Although this criterion is an inherent part of mediation, it is often not considered in social psychological research. One corollary of the timing criterion is that stable personality variables can never be mediators—unless we are talking about stable personality change in the context of longitudinal studies. It is important to consider that the conceptual timing criterion does not mean that *M* has to be necessarily *measured* before *Y* in an experiment. As long as the pure fact that *Y* is measured will not influence *M*, it is not problematic to measure *Y* before measuring *M*. This fact is often misunderstood by researchers who argue that in order to show a "true" effect of *M* on *Y* empirically, *M* necessarily has to be measured before *Y*.

5.2. Ways to improve the quality of inferences about the causal order

While we have shown that inferences about the causal order cannot be improved statistically by the RMT approach, this does not imply that social psychologists have no means to obtain better evidence for assumed mediation processes. On the one hand, there are statistical approaches to adequately control for biases (see above). On the other hand, to obtain further evidence about the causal order between the mediator and the dependent variable, social psychologists may want to change the research designs they typically use. There are several alternative experimental designs which can be grouped into three clusters.

First, there are approaches that include an additional experimental condition. In the *blockage design* (see MacKinnon, 2008; MacKinnon, Kisbu-Sakarya, & Gottschall, 2014), the experimental factor has a further level in which the manipulation of the independent variable is shaped in a way that blocks and removes the assumed mediator route. Consequently, the hypothesized mediation process can only operate in the condition in which it is not blocked. Evidence for the expected process is given when the independent variable has an effect only in this condition. To illustrate this design with an example, an experimental intervention is hypothesized to promote prosocial behavior by initiating cognitive processes. In the condition where the mediator is blocked, participants have to complete additional tasks so that cognitive load blocks the assumed pathway. When the related *enhancement design* (see MacKinnon, 2008; MacKinnon et al., 2014) is used, some participants are assigned to a condition in which the independent variable is manipulated in a way that is assumed to also intensify the value of the hypothesized mediator. Referring to the example just given, such a condition could additionally foster the expected cognitive processes. In accordance with the hypothesized process, the effect should be larger in this enhanced condition. While these modified designs are not without problems (see below) and not always feasible, they are worth considering.

The second approach is comparable to the previous one, but considers the mediator as a separate experimental factor. The idea is to manipulate whether the assumed process can or cannot run and to test the interaction between this variable and the independent variable. Support for the assumed mediation process is given when there is a simple main effect of the independent variable only in the condition in which the process is not interrupted. Accordingly, such a procedure is referred to as *moderation-of-process design* (Spencer et al., 2005) or *testing-a-process-hypothesis-by-an-interaction strategy* (TPIS, Jacoby & Sassenberg, 2011). Applications with, for instance, manipulated cognitive load (e.g., Galinsky & Moskowitz, 2007) or norm salience (e.g., Jonas et al., 2008) as mediators can be found. While this approach offers a useful option for those cases for which the intermediate stage of the assumed process can be manipulated independently from the manipulation of the independent variable, it—like the blockage and enhancement design—gives no concrete information about the specific causal links between the independent variable and the mediator as well as between the mediator and the dependent variable. In addition, the results obtained with these designs can be misleading. Sometimes they support the favored model although the specified mediation hypothesis does not apply. The variable assumed to be the mediator can be in fact a moderator (e.g., cognitive load) or just another variable associated with the dependent variable or might represent a sub-aspect of the construct assumed to be the dependent variable (Fiedler et al., 2011).

Third, Spencer et al. (2005) suggested *experimental-causal-chaining* (also called *double randomization design*; MacKinnon, 2008) as an alternative to the previous approaches. The idea is to conduct two separate experiments. In experiment 1, the independent variable is manipulated and its causal effect on the assumed mediator is tested empirically. Subsequently, in the second experiment, the mediator is manipulated and its causal effect on the DV is tested. Evidence for the assumed mediation process is given when both effects are statistically significant. This option, however, requires that the construct assumed to be the mediator can be measured as well as adequately manipulated in a similar meaning, which may not be plausible for all constructs. For instance, with reference to our example in the introduction, it is difficult to imagine that moral outrage can be manipulated in a way that is sufficiently comparable to measuring it. Generally, in this design it is conceptually often questionable whether the same aspect of the mediator construct is involved in both parts of the chain like it is stated in the mediation hypothesis. In addition, while this strategy enables appropriate conclusions about the separate causal links between the independent variable and the mediator as well as between the mediator and the dependent variable, it—like the other modified designs—gives no information on the value of the indirect effect and it does not include a comprehensive test of the mediation process $X \rightarrow M \rightarrow Y$ in a single study. Related to this, there is no information about the direct effect of X on Y. Hence, although resource demanding, the enrichment of the experimental-causal-chaining design with a third experiment in which, as usual, X is manipulated and M and Y are measured would be an informative strategy. Ideally, a mediation effect should be conceptually replicated with different designs (and different operationalizations).

6. Conclusion

The RMT approach to mediation analyses in which the mediator is measured (rather than manipulated) is generally not useful. Especially when the mediator is less reliable than the dependent variable, the RMT approach is likely to suggest wrong conclusions about the "correct" causal model. This does not mean that statistical tests of mediation hypotheses with data obtained from a typical social psychological experiment are useless or not worth being conducted. However, they cannot provide unequivocal evidence for the hypothesized causal order. For this, alternative research designs should be considered. Although none of the available approaches is a panacea, they offer possibilities that may be more or less reasonable given the specific research question. The repeated test of the assumed causal process with various strategies could help the "true" indirect effect to stand up.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2016.05.002.

References

Baron, R. M., & Kenny, D. A. (1986). Moderator-mediator variables distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Brown, R. L. (1997). Assessing specific mediational effects in complex theoretical models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*, 142–156.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550–558.

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables. Bootstrapping with structural equation models. *Organizational Research Methods*, *11*, 296–325.

Danner, D., Hagemann, D., & Fiedler, K. (2015). Mediation analysis with structural equation models: Combining theory, design, and statistics. *European Journal of Social Psychology*, *45*, 460–481.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*, 1231–1236.

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, *47*, 61–87.

Galinsky, A. D., & Moskowitz, G. B. (2007). Further ironies of suppression: Stereotype and counterstereotype accessibility following suppression. *Journal of Experimental Social Psychology*, *43*, 833–841.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927.

Hoyle, R. H., & Robinson, J. I. (2004). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 213–233). Thousand Oaks, CA: Sage Publications.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology*, *62*, 429–440.

Hyman, H. (1955). *Survey design and analysis: Principles, cases and procedures*. Glencoe, IL: The Free Press.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334.

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*, 180–190.

Jonas, E., Martens, A., Niesta Kayser, D., Fritsche, I., Sullivan, D., & Greenberg, J. (2008). Focus theory of normative conduct and terror-management theory: The interactive impact of mortality salience and norm salience on social judgment. *Journal of Personality and Social Psychology*, *95*, 1239–1251.

Jose, P. E. (2013). *Doing statistical mediation and moderation*. New York: Guilford Press.

Judd, C. M., & Sadler, M. S. (2008). The analysis of correlational data. In M. C. Roberts, & S. S. Ilardi (Eds.), *Handbook of research methods in clinical psychology* (pp. 115–137). Malden, MA: Blackwell.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum Associates.

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, *19*, 30–43.

MacKinnon, D. P., Kisbu-Sakarya, Y., & Gottschall, A. C. (2014). Developments in mediation analysis. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods. Statistical analysis, Vol. 2*. (pp. 338–360). New York: Oxford University Press.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, K., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 287–312.

Pearl, J. (2011). The causal mediation formula—A guide to the assessment of pathways and mechanisms. *Prevention Science*, *13*, 426–436.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, *19*, 459–481.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, *5*(6), 359–371.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290–312.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective in examining psychological process than mediational analyses. *Journal of Personality and Social Psychology*, *89*, 845–851.

Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, *37*, 235–246.

Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Journal of Basic and Applied Social Psychology*, *37*, 226–234.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*, 137–150.