



Reports

What mediation analysis can (not) do ☆

Klaus Fiedler ^{a,*}, Malte Schott ^a, Thorsten Meiser ^b^a University of Heidelberg, Germany^b University of Mannheim, Germany

ARTICLE INFO

Article history:

Received 28 April 2011

Available online 14 May 2011

Keywords:

Causal model

Spurious mediator

Sobel test

Attitude change

ABSTRACT

The present article is concerned with a common misunderstanding in the interpretation of statistical mediation analyses. These procedures can be sensibly used to examine the degree to which a third variable (Z) accounts for the influence of an independent (X) on a dependent variable (Y) conditional on the assumption that Z actually is a mediator. However, conversely, a significant mediation analysis result does not prove that Z is a mediator. This obvious but often neglected insight is substantiated in a simulation study. Using different causal models for generating Z (genuine mediator, spurious mediator, correlate of the dependent measure, manipulation check) it is shown that significant mediation tests do not allow researchers to identify unique mediators, or to distinguish between alternative causal models. This basic insight, although well understood by experts in statistics, is persistently ignored in the empirical literature and in the reviewing process of even the most selective journals.

© 2011 Published by Elsevier Inc.

As a prominent research aim is to understand the processes that underlie empirical phenomena, a key methodological concept is mediation. For empirical findings to gain real impact and to be published in a major journal, researchers should not merely describe the relationship between independent and dependent variable but also try to explain that relation in terms of mediating processes.

Mediation analysis (Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) is therefore considered an important research tool; it "... is now almost mandatory for new social-psychology manuscripts" (Bullock, Green & Ha, 2010, p. 550). In a nutshell, mediation analysis (MA) is a statistical procedure to test whether the effect of an independent variable X on a dependent variable Y (i.e., $X \rightarrow Y$) is at least partly explained by a chain of effects of the independent variable on an intervening mediator variable Z and of the intervening variable on the dependent variable (i.e., $X \rightarrow Z \rightarrow Y$).

In this article, we point out a basic misunderstanding about what MA can do and what it cannot do. We neither want to argue that the state-of-the-art statistical procedures used for MA are flawed or biased (cf. Bullock et al., 2010), nor do we postulate that scientists should refrain from MA. There can be no doubt that clarifying mediation is at the heart of high-quality research, and that the various statistical instruments developed for MA are appropriate if their stochastic and metric preconditions are met (see MacKinnon et al., 2002).

Our note pertains, rather, to a major category mistake concerning the theoretical insights that can be gained from such statistical analyses. That fundamental mistake consists in the widely shared belief that MA

can actually find out a mediator, or infer whether a particular variable is a unique mediator, or that significant mediation tests provide cogent evidence for the causal role played by the focal variable suggested in the preferred mediation model.

Such inferences are unwarranted and should be excluded from logically sound theoretical arguments. What MA can do is testing the significance, and maybe the effect size of a hypothetical mediator, assuming it is the actual mediator. However, MA is mute about the viability of the premise that the assumed intervening variable truly is a mediator. MA does not even allow for probabilistic inferences about the likelihood that the focal variable is a mediator as long as we do not know the likelihood distribution of all other potential mediators and alternative causal models of the relation between the independent, the dependent and the intervening variable. Although this insight is certainly not new (e.g., MacKinnon, Krull & Lockwood, 2000), we believe that these limitations are not sufficiently well-articulated in social psychology. Researchers, reviewers, and editors of leading journals take it for granted that mediators can be identified statistically. The literature on mediation analysis offers sophisticated insights into statistical procedures for estimating the parameters of given mediation models. The present critique, in contrast, is not concerned with any problems of statistical estimation but only with the logic of theoretical inferences that can be drawn from MA in the context of scientific discovery.

In the next section, we point out that a mediation model is only one of many possible causal models that can be used to describe a set of observed correlations or covariances, and that models capturing different theoretical assumptions may not be distinguishable statistically. In a following section, we present a simulation study to demonstrate that a variety of different causal data-generating processes can yield observed correlation patterns which spuriously resemble that of a true mediation process.

☆ Helpful comments by Daniel Danner, Joachim Krueger, Malte Friese, Susanne Beier and Laura DeMolière are gratefully acknowledged.

* Corresponding author.

E-mail address: kf@psychologie.uni-heidelberg.de (K. Fiedler).

Mediation analysis of a given set of correlations

We respectfully quote from well-done studies that use MA in intelligent ways but nevertheless exemplify the common misinterpretation of MA findings. With regard to long-term consequences of early attachment styles, Simpson, Collins, Tran, and Haydon (2007) conclude "... the current tests of the double-mediation hypothesis substantiate the contention that qualities of early caregiving are carried forward by the salient relationships of successive developmental periods" [p. 364]. With regard to the mediation of the impact of social contact on prejudice, Turner, Hewstone, and Voci (2007) reason that "self-disclosure improved explicit outgroup attitude via empathy, importance of contact, and intergroup trust" [p. 369] and that "we confirmed intergroup anxiety to be a mediator of the effect of direct and extended contact" [p. 382]. In a recent persuasion study, Tormala, Falces, Briñol, and Petty (2007) conclude: "unrequested cognitions played a mediating role in the ease of retrieval effect on judgment" [p. 143].

To examine the unwarranted inferences from MA more closely, let us refer to one prominent topic of social-psychological research, the elaboration of arguments in persuasion experiments (Meyers-Levy & Maheswaran, 1992; Tormala et al., 2007). In numerous experiments inspired by the elaboration likelihood model (Petty & Cacioppo, 1986), a common assumption is that attitude change via the central route (i.e., given sufficient cognitive resources) depends on the differential amount of supporting minus opposing cognitive responses in the recipient, as assessed in a thought-listing task. This measure is conceived as the crucial mediator variable. The impact of attitude quality (independent variable X) on attitude change (dependent variable Y) is supposed to be mediated by the recipients' cognitive responses (Z), as illustrated in Fig. 1. For a statistical test of this mediation hypothesis, researchers would typically follow the rules suggested by Baron and Kenny (1986), showing first that X is not only related to Y , but X is also related to Z and Z in turn to Y . By partialling out the third variable Z , it is then possible to test the hypothesis (e.g., via Sobel test) that the residual impact of X on Y is eliminated or reduced when the indirect path via Z is ruled out.

No doubt, such a test is clearly warranted and actually mandatory, because if Z is indeed a mediator, a logically sound implication is that the correlation between X and Y must be reduced when Z is partialled out. However, a typical problem with logical implications is that if-then statements cannot be reversed. If controlling for Z reduces the correlation between X and Y , this by no means implies that Z must be a mediator. No correlation statistics can prove that an alleged mediator is causally involved in the production of an effect. There is always the possibility that other potential mediators provide alternative explanations. Moreover, what appeared to be a mediator may actually play a different causal role (e.g., MacKinnon et al., 2000; Stelzl, 1986). Statistical analyses that focus only on one or a few selected mediators, whilst

neglecting countless other variables, can hardly contribute to identifying the true mediator. Statistical tests are always conditional on the premise: If a mediator is at work, is its impact significant? They cannot tell us whether a given variable is a mediator, because they cannot rule out that many other causal models provide an equivalent or even better account.

Let us again use the role of thought listing in persuasion experiments to illustrate this case. We can impose different causal interpretations on the tri-variate relationship between X , Y , and Z , reflecting completely different cognitive-process assumptions. Four possibilities are presented in Fig. 2. The first possibility (diagram a) is that Z is indeed a genuine mediator between X and Y . That is, strong arguments may actually elicit supportive cognitive responses in the recipient which may in turn cause the resulting attitude change.

Secondly, it is also possible that the third variable Z , as it was measured in the study, may only be a *spurious mediator*, that is, a correlate of another variable Z' , which is the real causally effective mediator (see diagram b). As Z is correlated with Z' , it can mimic the entire pattern of correlations between all three variables. Nevertheless, in spite of their substantial correlation, the psychological interpretation of Z' may be fundamentally different from Z . For example, the real mediator of the attitude influence may not be the number of supportive cognitive responses to the persuasive arguments but the sympathy for or identification with the communicator. Psychologically, such a mediator would be fundamentally different from cognitive responses to arguments, although it may also be manifested in a thought-listing task. This example highlights the general problem that misspecifications of the causal model, including omissions of true mediators, may lead to biased results and severe misinterpretations (Judd & Kenny, 1981, pp. 607f.).

A third possibility is that Z is not a causal mediator but simply a correlate of the dependent variable (diagram c), that is, another reflection of the resulting attitude. Z (i.e., supporting thoughts vs. counter-thoughts) may just represent an alternative measure of the attitude change induced by the difficulty treatment. This is particularly plausible when the thought-listing measure Z was assessed after the dependent measure Y , as is often the case in pertinent experiments. However, even when cognitive responses are assessed online, during message encoding, they may be conceived as an immediate manifestation of the dependent variable, attitude change. As an influence may arise exerted quickly, it may affect all attitude-relevant inferences from the beginning, including declarative measures (i.e., verbal and

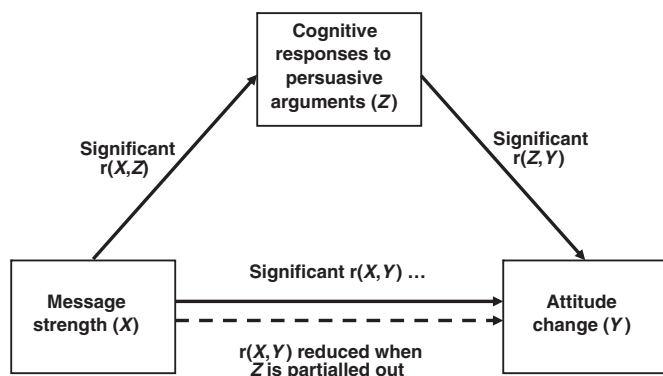


Fig. 1. Illustration of a mediation analysis in the context of persuasion research: The impact of message strength (X) on attitude change (Y) is supposed to be mediated by the recipients' cognitive responses (Z).

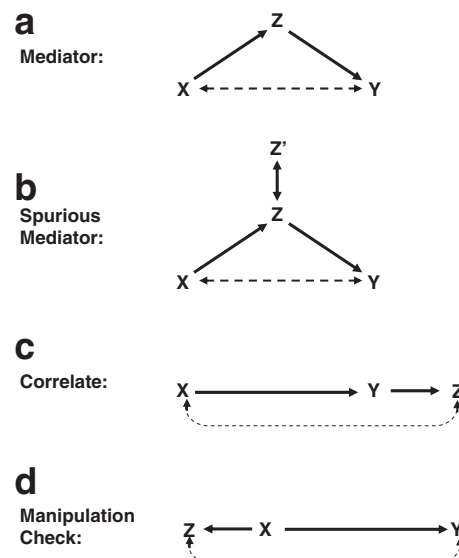


Fig. 2. Different causal models for the interpretation of the same intercorrelations between three variables, X , Y , and Z .

introspective thoughts) as well as procedural (cognitive or physiological) measures.

Assuming that Z is a correlate of Y can also explain the joint dependence of Z and Y on X : If both Z and Y reflect attitude change and if attitudinal effects in general reflect difficulty, X , then both Y and Z can be expected to correlate with X . The model that specifies Z as a correlate of Y therefore appears equally compatible with the data as the mediator model in diagram a. In fact, these two models are formally equivalent and thus indistinguishable on merely statistical grounds. Whether the underlying processes of the correlate model can actually produce a correlation pattern that mimics mediation is an empirical and numerical question that will be addressed in the simulation study reported below.

Analogous to the assumption that Z is a reflection of the dependent measure Y , the fourth and last model in diagram d assumes that Z is a second measure of the independent variable X . Let us call this role played by Z a manipulation check. That is, Z may be just another measure of argument strength. Supportive cognitive responses may indicate strong arguments whereas rejecting responses may indicate weak arguments. Indeed, the way in which strong and weak arguments are selected in pretesting often seems to follow this rationale. If, however, Z is but a manipulation check or a reflection of the construct behind X , then it makes perfect sense that Z like X also correlates with Y .

However, notably, although the manipulation check model can also assimilate the given correlation pattern, it differs from the other models in that it will not pass a MA test. If Z and Y are independent effects of a

common cause X , then controlling for Z should not affect the statistical relationship between X and Y , as will be explained below.

In any case, most empirical studies allow for many roles played by an alleged mediator Z , and for many alternative mediators Z' , Z'' , Z''' etc. to be ruled out. The truism that every empirical correlation can be in principle explained alternatively by a third variable applies to mediator variables as well. Moreover, the multitude of alternative explanations includes a variety of possible combinations of different causal models that may be jointly at work.

A simulation study

So far, we have only discussed the possibility that different causal models are equally compatible with a given pattern of correlations or, covariances and variances. To illustrate the crucial point that essentially different underlying data-generating processes may mimic an apparent mediator, we ran a simulation study. Under the constraints of the causal models in Fig. 2 (i.e., genuine mediator, spurious mediator, correlate, manipulation check), we generated normally distributed random variables X , Y , Z such that specific assumptions about their pairwise correlations $r(X,Y)$, $r(X,Z)$, $r(Y,Z)$ were met. These data were then subjected to MA, according to Baron and Kenny (1986), using the Sobel test statistic as a convenient index. One hundred replications were run for each combination of the fixed correlation parameters within each causal model. Means and standard deviations of the resulting Sobel z test statistics are reported below in Table 1 and Figs. 3 and 4.

Table 1

Mean and standard deviation (SD) of mediation analysis results (Sobel z) as a function of different causal models and two sample sizes (across 100 simulation trials).

<i>n</i> = 100						<i>n</i> = 200						
<i>r</i> _{xy}	<i>r</i> _{xz}	<i>r</i> _{zy}	<i>r</i> _{partial}	Sobel <i>z</i>		<i>r</i> _{xy}	<i>r</i> _{xz}	<i>r</i> _{zy}	<i>r</i> _{partial}	Sobel <i>z</i>		
				Mean	SD					Mean	SD	
0.08	0.30	0.30	−0.01	2.15	0.12	0.09	0.30	0.30	0.00	3.04	0.11	Mediator
0.16	0.30	0.50	0.01	2.70	0.04	0.15	0.30	0.50	0.00	3.84	0.04	
0.21	0.30	0.70	−0.01	2.95	0.01	0.22	0.30	0.70	0.01	4.19	0.01	
0.14	0.50	0.30	−0.01	2.45	0.41	0.16	0.50	0.30	0.01	3.38	0.32	
0.23	0.50	0.50	−0.02	3.76	0.20	0.24	0.50	0.50	−0.01	5.33	0.20	
0.36	0.50	0.70	0.01	4.71	0.09	0.35	0.50	0.70	0.00	6.71	0.09	Spurious mediator
0.21	0.70	0.30	0.00	2.15	0.61	0.22	0.70	0.30	0.01	2.99	0.60	
0.35	0.70	0.50	0.00	3.72	0.53	0.35	0.70	0.50	0.01	5.27	0.55	
0.49	0.70	0.70	−0.01	5.64	0.35	0.49	0.70	0.70	0.00	7.99	0.36	
0.07	0.30	0.21	0.01	1.64	0.39	0.07	0.30	0.21	0.01	2.37	0.46	
0.11	0.30	0.35	0.01	2.30	0.29	0.11	0.30	0.35	0.00	3.29	0.27	Correlate
0.16	0.30	0.49	0.01	2.67	0.10	0.16	0.30	0.49	0.01	3.80	0.10	
0.12	0.50	0.21	0.02	1.65	0.62	0.11	0.50	0.21	0.00	2.40	0.60	
0.18	0.50	0.34	0.01	2.70	0.51	0.17	0.50	0.35	0.00	3.95	0.57	
0.24	0.50	0.49	0.00	3.63	0.39	0.26	0.50	0.49	0.01	5.21	0.41	
0.16	0.70	0.21	0.01	1.42	0.72	0.15	0.70	0.20	0.01	1.91	0.80	Manipulation check
0.25	0.70	0.36	0.00	2.60	0.81	0.26	0.70	0.36	0.01	3.56	0.80	
0.34	0.70	0.48	0.01	3.54	0.79	0.35	0.70	0.49	0.01	5.17	0.65	
0.30	0.10	0.30	0.29	0.82	0.75	0.30	0.09	0.30	0.29	1.16	0.66	
0.30	0.15	0.50	0.26	1.44	0.84	0.30	0.15	0.50	0.27	1.98	0.75	
0.30	0.20	0.70	0.23	2.00	0.67	0.30	0.21	0.70	0.22	3.00	0.68	Cluster
0.50	0.15	0.30	0.48	1.15	0.45	0.50	0.15	0.30	0.48	1.77	0.43	
0.50	0.25	0.50	0.45	2.22	0.49	0.50	0.25	0.50	0.45	3.15	0.48	
0.50	0.35	0.70	0.38	3.33	0.58	0.50	0.35	0.70	0.38	4.75	0.54	
0.70	0.21	0.30	0.68	1.33	0.20	0.70	0.20	0.30	0.69	1.98	0.28	
0.70	0.36	0.50	0.65	2.61	0.21	0.70	0.35	0.50	0.65	3.79	0.15	Cluster
0.70	0.48	0.70	0.58	4.26	0.16	0.70	0.49	0.70	0.57	6.11	0.20	
0.30	0.30	0.10	0.29	0.12	0.83	0.30	0.30	0.09	0.29	0.01	1.02	
0.30	0.50	0.15	0.27	−0.03	1.00	0.30	0.50	0.14	0.27	−0.13	0.92	
0.30	0.70	0.22	0.21	0.16	0.91	0.30	0.70	0.20	0.23	−0.18	1.06	
0.50	0.30	0.15	0.48	0.00	0.97	0.50	0.30	0.15	0.48	0.01	0.90	Cluster
0.50	0.50	0.24	0.45	−0.08	1.03	0.50	0.50	0.26	0.44	0.12	0.99	
0.50	0.70	0.33	0.40	−0.32	0.87	0.50	0.70	0.34	0.39	−0.12	0.80	
0.70	0.30	0.21	0.69	−0.03	0.84	0.70	0.30	0.22	0.68	0.20	1.03	
0.70	0.50	0.34	0.65	−0.18	0.94	0.70	0.50	0.35	0.65	0.04	1.03	
0.70	0.70	0.49	0.57	0.02	1.03	0.70	0.70	0.49	0.57	0.03	1.02	Cluster
0.30	0.30	0.30	0.23	1.89	0.03	0.30	0.30	0.30	0.22	2.70	0.08	
0.50	0.50	0.50	0.33	3.01	0.07	0.50	0.50	0.50	0.33	4.28	0.08	
0.70	0.70	0.70	0.41	4.08	0.07	0.70	0.70	0.70	0.41	5.83	0.11	Cluster

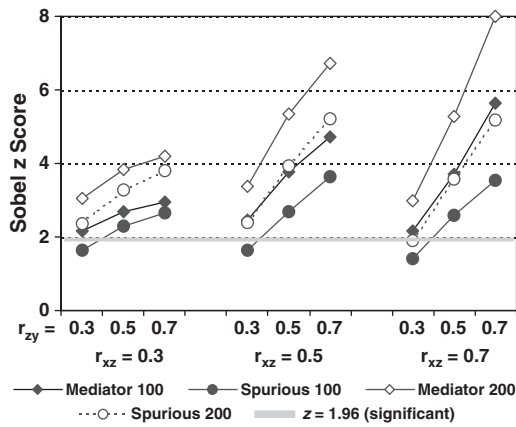


Fig. 3. Average Sobel z score obtained in mediational analysis for genuine mediators and spurious mediators as a function of sample size ($n = 100$ vs. 200) and different pairwise correlations between independent variable (X), dependent variable (Y) and hypothetical mediator (Z).

For the simulation, normally distributed random variables X , Y , and Z were generated according to the following rules. In the *genuine-mediation condition*, Z was first generated to bear a substantial correlation $r(X,Z)$ to the independent variable X . Then the dependent variable Y was generated to correlate at the level of $r(Y,Z)$ with the mediator Z . The resulting correlation between independent and dependent variable, $r(X,Y)$, was left open to vary as a sole function of the causal path from X to Z to Y .

In the *spurious-mediator condition*, we first established a genuine mediator relation, calling the mediator Z' . We then generated Z to bear a strong correlation, $r(Z,Z') = 0.7$ with the genuine mediator Z' .

In the *correlate condition*, Y was first generated to correlate at $r(X,Y)$ with X , and Z was then generated to correlate with Y , at the level $r(Y,Z)$. No further constraints were imposed on the remaining relation $r(X,Z)$.

Finally, in the *manipulation-check condition*, the $r(X,Y)$ relation was also generated first, and Z was constrained to bear a specific relation $r(X,Z)$ to the independent variable X , whereas $r(Y,Z)$ was unconstrained.

We restrict our simulation to these four generation models that we suspect to represent some of the most common misinterpretations of covariates. There are of course other models that might be simulated and different ways of generating the present models. We particularly refrain from more refined models that distinguish latent and measured variables, simply because the logical insights to be gained

here can be illustrated with a small set of simple models that, for the sake of simplicity, equate observed and latent variables.

The same procedure was repeated for different sample sizes ($n = 100$ and 200) and different patterns of correlation. For convenience, we let all generated correlations take on three different values, 0.3 , 0.5 , and 0.7 , supposed to represent weak, strong and very strong relations, respectively. In addition to the Sobel test measure z , we also monitored the values of those emergent correlations that were not constrained within certain causal-model conditions (e.g., the resulting value of $r(X,Z)$ given the correlate model).

Detailed numerical results for both samples sizes ($n = 100$ and 200) are given in Table 1. The first block of nine rows provides the data for the genuine mediator condition; the three subsequent blocks refer to the spurious-mediator, correlate, and manipulation-check conditions, respectively. (The last block of three rows, labeled “Cluster”, will be introduced in the final discussion.) From the first three columns it is evident that within each block, one of the three correlations is allowed to vary freely as a function of the other two correlations, which are set to values of 0.30 , 0.50 , or 0.70 . Marked in *italics*, the free-floating correlation is $r(X,Y)$ for mediators and spurious mediators, $r(X,Z)$ for the correlate model, and $r(Y,Z)$ for the manipulation-check model. The last two columns present the means and standard deviations of the Sobel z statistics obtained in all 100 simulation trials underlying each row. Fig. 3 summarizes the mean Sobel z values graphically, providing a convenient overall impression of significant MA results for genuine mediators as well as other causal models.

Obviously, the causal history of a significant Sobel test does not need to be a genuine mediation process. The Sobel test statistic z often exceeds the significance threshold ($z = 1.96$ for $\alpha = 0.05$), regardless of whether Z is actually a mediator or not. Moreover, the relatively small standard deviations in the right-most column of Table 1 suggest that these significant results can be expected to occur fairly often.

Consider the genuine-mediator condition first. Table 1 and Fig. 3 demonstrate that the statistical procedures do what they promise to do. When Z is a genuine mediator, evidence for mediation (manifested in z values exceeding the gray horizontal line in Fig. 3) increases regularly with increasing $r(X,Z)$ and also with increasing $r(Y,Z)$ levels. It is also evident from Table 1 that MA results increase from $n = 100$ to $n = 200$. Almost any reasonable pattern can be raised to significance by running a sufficiently large sample.

However, if Z is a spurious mediator, the resulting z are only slightly weaker, producing many significant results, if $r(Y,Z)$ and $r(X,Z)$ are not too small. The overall pattern is very similar to the genuine-mediator pattern, thus testifying to the well-known difficulty to distinguish between highly correlated variables (i.e., the true mediator Z' and its reflection Z).

A similar pattern arises when Z is simply a correlate of Y . If $r(Y,Z)$ is not too low – the premise of this model – most z values become significant, especially when n is large. Although the causal generation process in this condition only constrains the $r(Y,Z)$ correlation, a substantial relation between X and Y is sufficient to let $r(X,Z)$ also be strong enough in many cases to produce a systematic mimicry of Z being a mediator of the X – Y relation.

One noteworthy exception can be found in the manipulation-check model (cf. Fig. 4). When Y and Z are generated as independent (unconstrained) correlates of the same common cause X , there is no false evidence for spurious mediation effects. This clear-cut elimination of significant z results, even for very large samples, can indeed be expected on logical ground. Formally, the indirect effect of X on Y (via Z) is the product of two regression coefficients: (a) the regression coefficient of the intervening variable Z on the independent variable X , b_{ZX} , and (b) the regression of the dependent variable Y on Z in a multiple regression equation with X as additional predictor, $b_{YZ.X}$. In the manipulation-check scenario with X as a common cause of Z and Y , the relationship between the dependent variable Y and the alleged intervening variable Z is fully mediated by the independent variable X ,

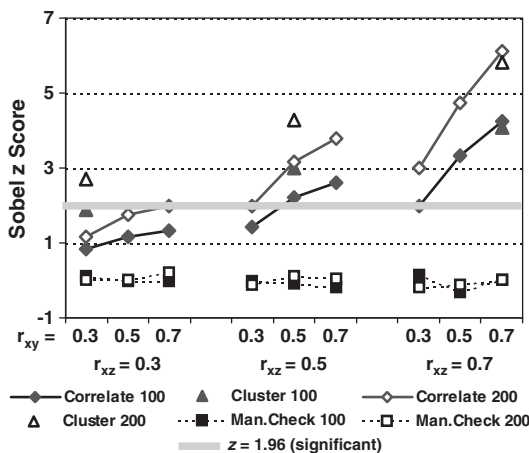


Fig. 4. Average Sobel z score obtained in mediational analysis for different causal models (correlate, manipulation check; cluster) as a function of sample size ($n = 100$ vs. 200) and different pairwise correlations between independent variable (X), dependent variable (Y) and hypothetical mediator (Z).

so that $b_{YZ \cdot X}$ approaches zero and, as a consequence, the product of b_{ZX} and $b_{YZ \cdot X}$ does not deviate from zero either. In the correlate scenario, in contrast, both b_{ZX} and $b_{YZ \cdot X}$ differ from zero for any substantial values of r_{XY} and r_{ZY} , thus mimicking significant mediation for sufficiently large samples.

Merits and perils of mediation analysis

Our critique of premature mediation inferences can be summarized in a straightforward message: the statistical rationale of MA, as stated by Baron and Kenny (1986) and MacKinnon et al. (2002), must be understood within a one-sided if-then logic. If it is true that Z is a mediator of the impact of X on Y , then MA can be used to test the statistical significance of this mediator model. However, backward inference is logically unwarranted. If a statistical test of Z happens to be significant, this cannot be taken as cogent evidence that Z is truly a causally effective mediator (or even a partial mediator) of the X – Y effect. Like any logical implication, “if Z is mediator, then it will be significant in a mediator analysis” is not symmetrical, simply because $X \rightarrow Z \rightarrow Y$ is not the only causal path leading to Y .

Behavioral phenomena are typically determined by multiple causes and several sensible levels of explanation. If one hypothetical mediator test proves significant, this does not mean that the hypothetical mediator is causally effective. Strictly speaking, it does not even imply an increase in the relative likelihood that the focal mediator is at work, compared to the likelihood of alternative mediators. Why should a significant Sobel test increase the likelihood that it reflects a genuine mediation effect more than the likelihood that it reflects a correlate of the dependent measure, attitude change, or an alternative mediator? For inferences from MA to causal processes to be possible, it would be necessary to specify the likelihood functions of all possible causal models. As such information is virtually never available, there can be no backward inference from a significant MA test to the validity or truth of causal assumptions.

What the reported simulations have shown, rather, is that a significant MA statistic can arise for different reasons. If a variable Z is generated to be a genuine mediator, it will indeed prove significant. However, Z will also be often significant when it is not a mediator but just a correlate of the dependent variable, Y , or a correlate of the true mediator with a fundamentally different theoretical meaning.

Let us try to paraphrase this message in terms of the psychology of scientific discovery. A researcher may not have strong theoretical reasons to assume that Z (rather than many other variables, Z' , Z'' , etc.) is causally responsible for an effect. However, Z may be an easily assessable variable that is somehow related to the observed Y effect, or Z may be a popular construct that fits the current mainstream, or a confound of genuine mediators that are unknown and not included in a study or that are not amenable to simplistic measurement. Then, as we have seen, Z may nevertheless mimic a mediator and survive a significance test.

Let us depict one more alternative model to highlight this scenario. This model is labeled “Cluster” in Table 1 and Fig. 4. Let us assume our researcher has simply gathered three redundant measures X , Y , and Z of the same latent construct. Their causal structure is completely unknown. Given such a diffuse cluster, the researcher has an easy time to find statistical support for any of the three variables as a mediator of the influence between the other two, in any direction, supporting the researcher’s favorite hypothesis. Thus, given three homogeneously correlated variables, MA may confirm any arbitrary mediation model.

To illustrate this diffuse case with the persuasion example, let us assume that argument strength (X), posterior attitude (Y), and the tendency to generate supportive cognitive responses (Z) all reflect the same latent construct, the attitude itself. To simulate this case, we ran MAs at three levels of homogeneous intercorrelations that do not place any constraints on the causal process, setting $r(X,Y) = r(X,Z) = r(Y,Z) = 0.3, 0.5$, and 0.7 , respectively. The triangles in Fig. 4 show that all z values but those for the smallest r at the smallest n level were significant. Apparently, drawing a sample of redundant variables from

a homogeneous set is sufficient to mimic a mediation effect, provided that the sample size is not too small.

We do not want to be misunderstood as conveying a purely skeptical message regarding the theoretical usefulness of MA. Our critique is rather meant to be constructive and built on the firm ground of what we have all along taught our students about correlation statistics. Like all correlation statistics, MA is not essentially weak or misleading. If it is correctly taken for what it is – namely, an elaborate methodology for testing correlational hypotheses – there is a lot to gain from MA. It offers sound procedures for testing the strength and significance of a selected causal model, or for comparing different causal models. However, crucially, like all correlation methods, it cannot identify causes or true mediator variables that play the leading causal role within a theoretical explanation. Whether a selected causal variable reflects a real cause or not cannot be determined statistically. This is a matter of clever theorizing, convergent validation, analytical conventions, and experimental tests of antecedence–consequence relations (Roberts & Pashler, 2000; Campbell & Fiske, 1959; Slovic, 1962).

Research by Fiedler, Walther, Freytag, and Stryczek (2002) illustrates this point with regard to the distinction of a mediator versus a moderator model. On each trial of a stimulus series, participants saw a cartoon of a male agent trying to mate a female partner using either a direct or an indirect verbal strategy. Across all trials one strategy (e.g., direct) was more successful, leading to more positive female responses. However, within both subsets of females, cool women and emotional women (symbolized by turquoise and red head color, respectively), the other (e.g., indirect) strategy was more successful. To solve this variant of Simpson’s (1951) paradox, participants had to find out that emotional women give generally more positive responses and that direct strategies mostly came along with emotional women.

Participants could either interpret women type as a moderator or as a mediator. If the difference of cool versus emotional women existed (visible in head color) as a moderating condition prior to the male utterance, the seeming advantage of a direct strategy can be discounted as merely due to the easier subset of women approached with a direct strategy. In contrast, if women types were framed as a mediator in that females took on their head color only after the male strategy, suggesting that a direct strategy (X) elicited an emotional reaction (Z) which in turn produced a positive response (Y), then one might reasonably continue to believe that direct strategies lead to more success (mediated by female emotion) than indirect strategies. Indeed, judgments of the success of direct strategies were higher in the mediator condition than in the moderator condition. Thus, the same correlation pattern was either interpreted as a mediation effect (if Z followed X in close contiguity) or as a moderator effect (if Z was predetermined). Correlation statistics contribute nothing to disambiguate the causal model. That “correlations are an instrument of the devil” (Birnbaum, 1973) may sometimes be true, but only when researchers hold unwarranted expectations about correlations. All correlation analyses have a Protean quality; their answers are severely biased towards the questions that researchers are asking. They themselves cannot identify causes or crucial mediators that are not already presupposed in the research question. They cannot even demonstrate an increase in the posterior probability that a significant Z is actually a mediator.

The Kantian argument that a mediator is assumed a priori rather than found a posteriori should neither discourage researchers nor undermine the simple beauty of correlation statistics. They offer standardized ways of describing empirical relationships, a natural measure of agreement, and a generally applicable index of effect size. Moreover, correlation methods are sorely needed as basic statistical modules in spectral analysis, Fourier analysis, information theory, artificial intelligence, and in many other contexts.

Correlational analyses are not generally inferior to experimental analyses. They are indispensable to illuminate certain causal constellations that experimental designs would conceal. This is particularly the case for mediators, which are systematically ruled out by orthogonal

designs. For example, if the mediation assumption is true – which is still a sensible and theoretically inspiring option – a natural correlation between argument strength (X) and cognitive support (Z) carries over to an attitude change effect (Y). Forcing X and Z into the straight jacket of an orthogonal design would artificially undo the correlation that is crucial to understanding the psychological process leading from X to Z to Y .

Thus, there is no alternative to advanced correlational methods when it comes to analyzing and testing naturally occurring mediation effects. In order to substantiate the hypothesis that Z represents a mediator of the X – Y relation, a significant mediation test would afford a *necessary* condition. It is, however, not a *sufficient* condition for the identification of Z as a mediator that was not identified or assumed before. In other words, MA must not be misunderstood as the final link in a scientific inference process that leads to the a-posteriori identification of the true or ultimate or most important mediator. All that MA can do is checking on a necessary condition of a causal model that already exists on a-priori grounds.

Even when Z (e.g., cognitive responses) is a genuine mediator, the same relation can also be construed as mediated by Z' (e.g., sympathy-based motivational state) or by Z'' (e.g., suggestibility as a personal disposition) or by Z''' (e.g., underlying neural mechanism), and so forth. Which level of explanation is chosen depends on the researcher's embedding theory, her scientific discipline, the kind of planned interventions, and her intuitions about what constitutes a fertile and promising theoretical approach.

There is no a-priori reason to believe in the existence of a unique mediation model. To decide whether Z is a genuine mediator, as specified by the model $X \rightarrow Z \rightarrow Y$, researchers may reasonably try to manipulate mediators experimentally, though experimental mediation analysis can be problematic too (Bullock et al., 2010). For instance, if an influence of X on Y remains significant although Z is held constant or controlled statistically, then Z cannot be assumed to mediate the residual effect. Likewise, if the orthogonal manipulation of X (independent variable) and Z (mediator candidate) eliminates the $X \rightarrow Z$ correlation, a persisting influence of X on Y could be no longer due to Z as a mediator. These examples highlight, though, that such experimental tests will more likely lead to the falsification of a mediator hypothesis than to its verification – a truism uncovered by Popper (1959).

Conclusion

To summarize, even when alternative theoretical models are considered by a researcher, different theoretical assumptions concerning the role of intervening variables can be empirically indistinguishable because of mathematical model equivalence (Mackinnon et al., 2000; Stelzl, 1986). Therefore, the causal role of intervening variables cannot be determined by statistical analysis alone, but design features like experimental manipulation with random assign-

ment, longitudinal analysis of temporal antecedence, or other control mechanisms (e.g., Rubin, 2006) have to be employed to circumvent the limitations of correlational analysis and its interpretation. As is the case with any statistical hypothesis, empirical confirmation of a mediation hypothesis (i.e., decomposition of a total effect $X \rightarrow Y$ into a significant indirect effect $X \rightarrow Z \rightarrow Y$ and a reduced direct effect $X \rightarrow Y$ controlling for Z) does not imply that the underlying psychological hypothesis (i.e., causal chain $X \rightarrow Z \rightarrow Y$) is true. The statistical hypothesis that is tested in a mediation analysis may be confirmed although the empirical pattern of correlations is generated by other processes than those specified in the presumed theoretical model. Pointing out this basic insight was the purpose of the present article.

References

- Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Birnbaum, M. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, 79, 239–242.
- Bullock, J., Green, D., & Ha, S. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550–558.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Fiedler, K., Walther, E., Freytag, P., & Stryczek, E. (2002). Playing mating games in foreign cultures: A conceptual framework and an experimental paradigm for inductive trivariate inference. *Journal of Experimental Social Psychology*, 38, 14–30.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.
- Mackinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1, 173–181.
- Mackinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83–104.
- Meyers-Levy, J., & Maheswaran, D. (1992). When timing matters: The influence of temporal distance on consumers' affective and persuasive responses. *Journal of Consumer Research*, 19(3), 424–433.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
- Popper, K. (1959). *The logic of scientific discovery*. Oxford England: Basic Books.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358–367.
- Rubin, D. B. (2006). *Matched sampling for causal effects*. New York: Cambridge University Press.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, 13, 238–241.
- Simpson, J., Collins, W., Tran, S., & Haydon, K. (2007). Attachment and the experience and expression of emotions in romantic relationships: A developmental perspective. *Journal of Personality and Social Psychology*, 92, 355–367.
- Slovic, P. (1962). Convergent validation of risk taking measures. *The Journal of Abnormal and Social Psychology*, 65(1), 68–71.
- Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21, 309–331.
- Tormala, Z., Falces, C., Briñol, P., & Petty, R. (2007). Ease of retrieval effects in social judgment: The role of unrequested cognitions. *Journal of Personality and Social Psychology*, 93(2), 143–157.
- Turner, R., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, 9, 369–388.