Journal of Experimental Social Psychology 47 (2011) 1231–1236

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

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.

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).

0022-1031/$ – see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.jesp.2011.05.007

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

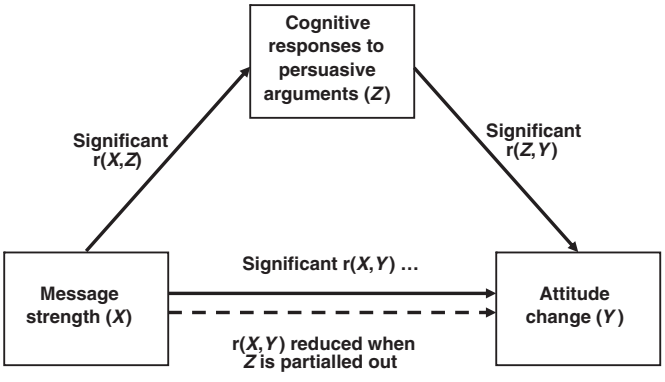

Fig. 1. Illustration of a mediation analysis in the context of persuasion research: The impact of message strength (*X*) on attitude change (*Y*) is supposed to be mediated by the recipients' cognitive responses (*Z*).

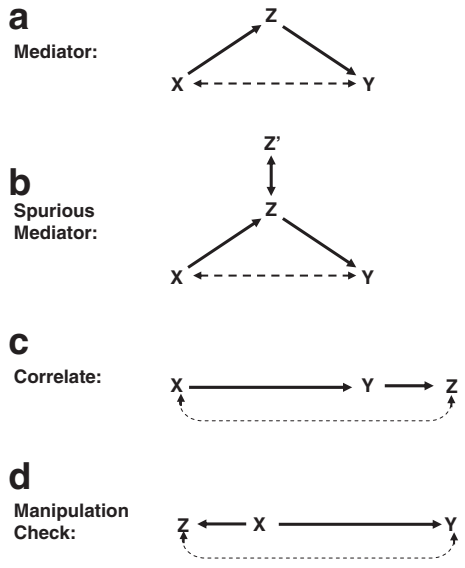

Fig. 2. Different causal models for the interpretation of the same intercorrelations between three variables, *X*, *Y*, and *Z*.