

Measuring evidence for mediation in the presence of measurement error

Arash Laghaie

Thomas Otter

October 21, 2021

Abstract

Mediation analysis empirically investigates the process underlying the effect of an experimental manipulation on a dependent variable of interest. In the simplest mediation setting, the experimental treatment can affect the dependent variable through the mediator (indirect effect) and/or directly (direct effect). Recent methodological advances made in the field of mediation analysis aim at developing statistically reliable estimates of the indirect effect of the treatment on the outcome. However, what appears to be an indirect effect through the mediator may reflect a data generating process *without* mediation, regardless of the statistical properties of the estimate. To overcome this indeterminacy where possible, we develop the insight that a statistically reliable *total* effect combined with strong evidence for conditional independence of treatment and outcome given the mediator is unequivocal evidence for mediation (as the underlying causal model generating the data) into an operational procedure. Our procedure combines Bayes factors as principled measures of the degree of support for conditional independence, i.e., the degree of support for a null hypothesis, with latent variable modeling to account for measurement error and discretization in a fully Bayesian framework. We illustrate how our approach facilitates stronger conclusions by re-analyzing a set of published mediation studies.

Keywords: mediation, causal model, identification, measurement, Bayes factor

INTRODUCTION

Mediation analysis is a tool to corroborate hypotheses about the process by which the experimental treatment brings about its effect on the dependent variable. Establishing mediation is fundamental to scientific research that seeks to understand "how the world works" in situations where the process connecting a manipulation to an outcome of interest is not directly observable. Understanding the process can be particularly important in marketing in situations where an action might fail to generate the desired outcome in the absence of the required mechanism. For example, in the context of targeting mobile ads, [Andrews et al. \(2016\)](#) show that individuals in crowded areas are more likely to respond to mobile ads because crowding invades individuals' privacy and causes them to immerse in their mobile phones and be exposed to ads. Without knowing the mechanism through which crowdedness affects ad response, a marketer might decide to develop an algorithm purely based on crowdedness that targets individuals in all high density areas including e.g., social events or live concerts. This would be futile because individuals seek social interactions at such events (or focus their attention on the performing artist at live concerts) and hence do not experience a privacy invasion to withdraw from. Put differently, the mechanism that channels the crowdedness effect into increased ad response in some crowded areas such as at e.g., a train station, does not exist in other crowded areas such as at social events or live concerts.

Moreover, mediation is at the heart of a structural understanding of the world that would otherwise require a new theory for every manipulation studied. For example, the observation that color contrast ([Thompson and Ince, 2013](#)), typicality ([Landwehr et al., 2011, 2013](#)), exposure ([Ferraro et al., 2008](#); [Landwehr et al., 2017](#)), or pronounceability ([Alter and Oppenheimer, 2008](#); [Song and Schwarz, 2009](#)) manipulations all bring about effects on a variety of dependent variables mediated by processing fluency, parsimoniously structures our causal understanding of the world ([Graf et al., 2018](#)). Relatedly, mediation defines subsets of variables that can be validly studied in isolation. For example, the understanding that the effect of variable production costs on demand is fully mediated by retail price allows for the exclusion of variable production costs from the price-demand model (and makes variable production costs a potential instrumental variable for price).

In marketing, in addition to its prevalence in the consumer behavior literature¹, mediation analysis has been used to study mechanisms across a wide range of topics: to understand advertising effects in general, e.g., the process by which advertising expenditures affect profit (Steenkamp and Fang, 2011), and in particular, e.g., the mechanisms by which word of mouth leads to persuasion (Packard and Berger, 2017), how favorable thoughts elicited by targeted versus personalized offers mediate the effect of personalization on ad response (Kramer et al., 2007), or how price promotions on future purchases in the form of gift cards are effectively overvalued through "double mental accounting" (Cheng and Cryder, 2018); in the context of product related decisions it is used to answer questions such as through what psychological channel mass customization affects willingness to pay (Franke et al., 2010), or how product variety affects brand choice (Berger et al., 2007); and in digital marketing in addition to our introductory example, to understand e.g., how availability of online product information increases offline purchase intentions (Bhargave et al., 2016).

In the context of corroborating hypotheses about unobserved processes such as mediation, it is useful to distinguish between uncertainty about the strength of unobserved connections that are beyond questioning per se, and uncertainty about the connective structure itself. Empirical testing of the strength of connections in a given, undisputed model is a standard statistical exercise that may require sophistication nevertheless. Resolving uncertainty about the connective structure, and ideally establishing a unique causal interpretation of the connective structure supported by the data, is usually more difficult and often beyond reach in the context of observed data.

The goal of mediation analysis in practice squarely falls into the latter class of problems. When conducting mediation analysis, researchers do not ask for measures of coefficients in a model they know a priori. Instead, researchers ask what model generated the data. To this end, this paper provides tools and a procedure aimed at measuring data-based evidence for mediation in the context of so called measurement-of-mediation designs with randomly assigned X and measured M and Y.

Our contribution is the synthesis of i) the causal identifiability of full mediation from observed variables (Glymour, 2001; Pearl, 2009), ii) Bayes factors as measures of *empirical evidence supporting* full mediation (Nuijten et al., 2015), and iii) explicit models

¹According to Pieters (2017), 71% of all experimental studies published in *Journal of Consumer Research* between years 2014-2016 contained at least one mediation analysis

of measurement error (Holmbeck, 1997; Bolger, 1998; Hoyle, 1999; Cole and Maxwell, 2003; MacKinnon, 2012) into a novel operational procedure. The user benefit from our synthesis is potentially strong empirical support for causal mediation as the only possible underlying data generating process. This includes an empirical confirmation of the assumed causal order of measured variables, i.e., empirical evidence against a reversed causal ordering of what is assumed to be the mediator and the dependent variable. As a byproduct, our procedure can establish the experimental manipulation, the mediator and the dependent variable as different causal entities relative to the competing hypotheses that measured variables are mere manipulation checks or multiple measures of one and the same dependent variable.

The remainder of this article is organized as follows: Section 2 characterizes the model discovery problem we address in this paper and Section 3 relates our contribution to extant literature. Section 4 develops the proposed procedure that operationalizes a causal identifiability result leveraging Bayesian model comparison (Bayes factors) and models of measurement error. Section 5 puts the procedure to practical use by re-analyzing recently published studies and illustrates its practicability and relevance. Finally, Section 6 summarizes our proposed approach in the form of a flowchart and a set of procedural steps to follow and discusses future research. Computer code implementing the developments in this paper is available online at [Github](#).

PROBLEM CHARACTERIZATION

Canonically, researchers seek support for the causal model depicted in the form of a directed acyclic graph (DAG) in Figure 1. In line with the vast majority of published applications, we will assume linear relationships between variables.²

²While this is not without loss of generality, the linear case is practically relevant as apparent from the mediation studies we re-analyze in Section 5. However, we later accommodate non-linear relationships between (linearly related) latent variables M and Y and their observed measures.

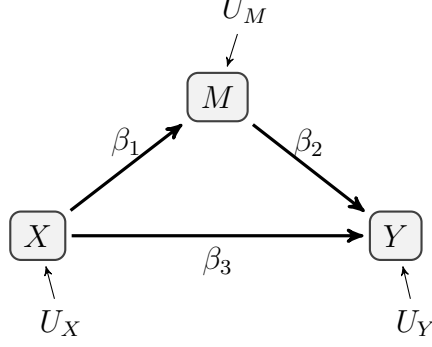


Figure 1: Mediation DAG

In this DAG, the experimentally manipulated, randomly assigned variable X exerts an effect on the dependent variable Y that consists of an indirect effect mediated by M , and a direct effect β_3 . Both M and Y are measured variables in the setting we study. Variables $\{U_M, U_Y\}$ are unobserved causes of variation in M and Y , respectively. Variable U_X symbolically summarizes experimental manipulation and random assignment of X (as a function of U_X , $X(U_X)$). As a consequence of random assignment, U_X and $\{U_M, U_Y\}$ are independent and the causal effect from X to M , β_1 , as well as the total causal effect from X to Y , $\alpha = \beta_3 + \beta_1 \cdot \beta_2$ can be consistently estimated.

Further, if this DAG is the data generating mechanism, it follows, first, that blocking the path through M by forcing M to take a particular value experimentally will make the effect from X to Y equal β_3 (will make the effect vanish in the case of $\beta_3 = 0$). This is the definition of mediation. Second, it follows that both β_2 and β_3 can be consistently estimated from observations of M and Y in an experiment manipulating X . This second implication is predicated on the independence between the disturbance terms U_M and U_Y that will only obtain if the assumed causal ordering of measured variables is correct.

However in practice, researchers often express considerable uncertainty regarding the causal ordering of measured variables (e.g., [Tonietto and Malkoc \(2016\)](#), see also the discussion in [Pieters \(2017\)](#)). Designs that measure the putative mediator only after measuring the putative dependent variable further complicate the definite causal ordering of measured variables based on prior knowledge (e.g., [Shaddy and Fishbach, 2017](#); [Sussman and O'brien, 2016](#)). Similarly problematic are unobserved background variables that render U_M and U_Y dependent even conditional on the correct causal ordering. As we illustrate below, these possibilities call into question the interpretation of results from standard mediation analysis as data-based evidence for mediation.

Historically, mediation analysis relied on a procedure popularized by [Baron and Kenny \(1986\)](#). The procedure fits the regression models in Equations 1-3, where α measures the *total causal* effect from X to Y and β_1 , the *causal* effect from X to M, when X is experimentally manipulated and thus randomly assigned. The coefficients β_2 , and β_3 in Equation 3 only have a causal interpretation when the error terms are subject to the assumptions expressed in Figure 1 and discussed above.

$$Y = \alpha_{0Y} + \alpha X + e_Y \quad (e_Y = \beta_2 U_M + U_Y) \quad (1)$$

$$M = \beta_{0M} + \beta_1 X + U_M \quad (2)$$

$$Y = \beta_{0Y} + \beta_2 M + \beta_3 X + U_Y \quad (3)$$

Thus, the correspondence between estimates obtained from fitting Equations 1-3 to data and the structural parameters in Figure 1 is uncertain, and the scope for learning about the actual data generating structure limited.

While results where the data credibly support a non-zero product $\beta_1 \cdot \beta_2$ when fitting Equations 1-3 are correctly interpreted as *consistent* with the causal model in Figure 1 these results, however, do *not* identify mediation as the data generating mechanism, and regardless of how much sophistication goes into assessing the statistical uncertainty in the estimate of the product $\beta_1 \cdot \beta_2$ (cf. [MacKinnon et al., 2002](#); [Shrout and Bolger, 2002](#); [MacKinnon et al., 2004](#); [Preacher and Hayes, 2004](#); [MacKinnon et al., 2007](#)). For an intuitive demonstration of this fact, we simulate data from four different underlying causal models to show how estimates of putative indirect effects can be misleading as evidence for mediation. We discuss existing approaches to overcome some of the indeterminacies illustrated here, and how we add to these approaches in the following section.

Case 1: We generate data (N=1000) from the model in Figure 2 with the following parameters (see Equations 1-3): $X \sim N(0, 1)$, $\beta_1 = 1$, $\beta_2 = 3$, $\beta_3 = 0$, $Var(U_M) = 1$, $Var(U_Y) = 1$. In this model the effect is fully mediated, as there is no direct effect from X to Y after controlling for M. Estimating Equations 2 & 3 yields, as expected, a significant indirect effect ($\widehat{\beta_1 \cdot \beta_2} = 2.99$) with a bootstrapped confidence interval of $95\%CI_{\beta_1 \cdot \beta_2} = [2.8, 3.2]$, that given knowledge of the data generating model in Figure 2

measures a causal effect that is in line with the data generating values.

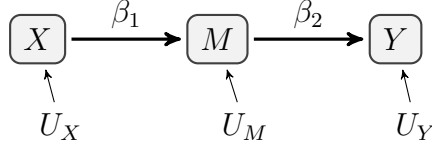


Figure 2: DAG of Case 1

Case 2: We generate data (N=1000) from the model in Figure 1 with the same parameter settings as above, but with a non-zero $\beta_3 = 1$. Both [Baron and Kenny's](#) and bootstrapped estimates of the indirect effect are significant ($\widehat{\beta_1 \cdot \beta_2} = 3.02$, bootstrapped $95\%CI_{\beta_1, \beta_2} = [2.8, 3.2]$) and again measure a causal indirect effect given the data generating model in Figure 1, and in line with the data generating values.

Case 3: Here the data (N=1000) are generated from the model in Figure 3. This model features a causal link between X and M. However, the dependence between M and Y in this model is purely correlational as a function of unobserved confounding variables. That is, there is no indirect or direct effect from X to Y. The data generating parameters are: $X \sim N(0, 1)$, $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = 0$, $Var(U_M) = 1$, $Var(U_Y) = 1$, $Cov(U_M, U_Y) = -0.5$. When fitting Equations 2 & 3 to this data, the estimate of (what could be) an indirect effect is significant: $\widehat{\beta_1 \cdot \beta_2} = -0.46$, bootstrapped $95\%CI_{\beta_1, \beta_2} = [-0.52, -0.40]$. Clearly, interpreting this estimate as evidence for (or a measure of) mediation is misleading in a context where one cannot rule out Figure 3 as the underlying causal structure based on prior knowledge in favor of Figure 1 or Figure 2.³

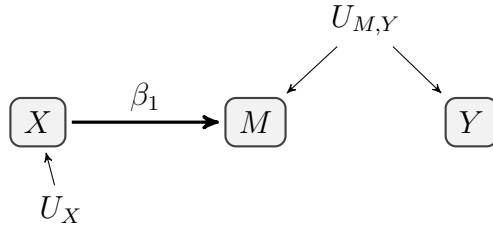


Figure 3: DAG of Case 3

Case 4: In this example the data are generated from the causal structure in Figure 2 using the exact same parameter values as in case 1 above. However, the analyst mistakes

³As an aside, the causal structure in Figure 3 implies a total effect from X to Y exactly equal to zero in increasing samples. The observation that standard mediation analysis will deliver estimates of what looks like a non-zero indirect effect (in addition to a non-zero direct effect) calls into question the general validity of continuing the search for mediators upon finding a zero total effect as advanced in [Zhao et al. \(2010\)](#).

what in reality is Y for M, and vice versa. The analyst thus estimates a model that corresponds to (mistakenly) assuming Figure 4 as the underlying causal structure. Here, again, there obviously does not exist a causal effect from the assumed mediator (that in reality is Y), to the assumed dependent variable (that in reality is M). However, standard mediation analysis (spuriously) infers a significant indirect effect: $\widehat{\beta_1 \cdot \beta_2} = 0.94$, Bootstrapped $95\%CI_{\beta_1, \beta_2} = [0.87, 1]$, as well as a direct effect (not reported here). Again, interpreting this estimate as evidence for (or a measure of) mediation is misleading because the true causal effect from X through Y to M is zero because of the lack of a causal connection from Y to M in Figure 2.

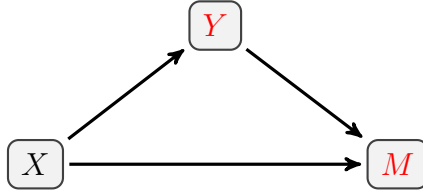


Figure 4: Misspecified DAG of Case 4

These illustrative simulations show that estimates of what may be an indirect effect can and should not be interpreted as evidence for, or measures of mediation without at least a discussion, or better, an empirical assessment of the underlying causal model. Note that this caveat cannot be overcome by investing in the statistical quality of what may or may not be an estimate of an indirect effect, e.g., using bootstrapping or Bayesian inference. Moreover, we show in the following section that even advanced extant approaches to causal mediation analysis at least depend on correct prior knowledge about which measured variable is M and which one is Y, an assessment that is often fraught with uncertainty in consumer behavior experiments.

RELATED LITERATURE

We organize the discussion of the literature related to our contribution using the classification scheme in Table 1. The first columns in Table 1 references a selection of papers representative of the respective approach to mediation analysis. Columns two to five in Table 1 pertain to assumptions that may be required for valid inference. The last column simply states if a particular approach measures evidence for conditional independence between randomly assigned X and what is taken to be a measure of Y, given what is taken to be a measure of M.

	Required assumptions				
	Asymptotic normality	No measurement error	No unobserved confounder ($U_{M,Y}$)	Causal direction b/w measured variables	Measures evidence for conditional independence
Baron and Kenny (1986)	Yes	Yes	Yes	Yes	x
Preacher and Hayes (2004)	relaxed	Yes	Yes	Yes	x
Muthén and Asparouhov (2015)	relaxed	relaxed	Yes	Yes	x
Imai et al. (2010) , Liu and Wang (2021)	relaxed	relaxed	relaxed	Yes	x
Zhang, Wedel, and Pieters (2009)	relaxed	relaxed	relaxed	Yes	x
The proposed procedure	relaxed	relaxed	relaxed	relaxed	✓

Table 1: A comparison of the existing approaches in terms of their assumptions for valid inference

Among the approaches listed in Table 1, only the original procedure by [Baron and Kenny \(1986\)](#) requires asymptotic normality of the product term $\beta_1 \cdot \beta_2$ for valid inference. This assumption has been relaxed subsequently using bootstrapping (e.g., [MacKinnon et al., 2002](#); [Shrout and Bolger, 2002](#); [MacKinnon et al., 2004](#); [Preacher and Hayes, 2004](#)) or Bayesian inference (e.g., [Zhang et al., 2009](#)). However, interpreting an estimate of $\beta_1 \cdot \beta_2$ as a consistent measure of a causal indirect effect requires further assumptions that pertain to structural aspects of the unobserved data generating model, i.e., a lack of measurement error in the mediator (column 3), the absence of unobserved confounders between measured variables (column 4), and the correct causal ordering of measured variables (column 5). Meeting this last requirement is not trivial in behavioral studies that often only measure what may be the causal mediator M after measuring what is assumed to be the dependent variable Y (e.g., [Shaddy and Fishbach, 2017](#); [Sussman and O’Brien, 2016](#)).

[Muthén and Asparouhov \(2015\)](#) relax the assumption of the absence of measurement error advocating latent variable modeling for mediation analysis (see also [Lee, 2007](#)). [Imai et al. \(2010\)](#) and [Liu and Wang \(2021\)](#) alleviate the stringency of the subjective bet on the absence of measurement error and that of unobserved confounders ($U_{M,Y}$) suggesting sensitivity analyses. Finally, [Zhang et al. \(2009\)](#) propose latent instrumental variable modeling to overcome both assumptions.

However, as can be seen from column 5 in Table 1, all extant approaches rely on prior knowledge about the causal ordering of measured variables. In other words, they require a definite a priori assessment of which measured variable is the mediator M and which is the dependent variable Y . The sensitivity analysis advocated in [Imai et al. \(2010\)](#) and [Liu and Wang \(2021\)](#) cannot distinguish between error correlations due to a reversed causal order or due to other reasons. The latent instrumental variable approach employed by [Zhang et al. \(2009\)](#) fails to identify the causal order because the instrument is latent. Online Appendix A provides numerical illustrations of how the sensitivity analysis originally proposed by [Imai et al. \(2010\)](#) and latent instrumental variable modeling ([Zhang et al., 2009](#)) will fail to detect a misspecified causal ordering of observed variables, i.e., the case where the researcher mistakes M for Y and vice-versa.

In contrast, the procedure we propose in this paper, and develop in detail in the following section, can unequivocally establish the causal chain from X through M to Y , and hence full causal mediation, as the data generating causal model. It is best viewed as a novel operationalization of an important result in the causality and model discovery literature (e.g., [Glymour, 2001](#); [Pearl, 2009](#)). In this context, we require a measure of data based support for conditional independence between X and Y given M . Our proposed solution to this challenge relies on Bayesian model comparisons. Hence, our inferential approach is fully Bayesian.

Bayesian inference, of course, has a history in mediation analyses (e.g., [Zhang et al., 2009](#); [Muthén and Asparouhov, 2012](#); [Dyachenko and Allenby, 2020](#); [Wedel and Dong, 2020](#)). However, to the best of our knowledge Bayesian inference has not been applied to the problem of model identification (causal discovery) in the context of mediation analysis. The main reasons for employing Bayesian inference for mediation analysis hitherto are i) better inference for the indirect effect by avoiding the appeal to asymptotic normality (e.g., [Zhang et al., 2009](#)), ii) optimal shrinkage to a hierarchical prior in studies with repeated measurements (e.g., [Wedel and Dong, 2020](#)), iii) computational reasons in the context of mediation models with latent variables and limited dependent observed variables (e.g., [Muthén and Asparouhov, 2012](#)), and iv) an attempt to model structural heterogeneity in mediation ([Dyachenko and Allenby, 2020](#)). While the latter seemingly provides indirect data-based evidence for the causal chain from X through M to Y , we explain and numerically illustrate in Appendix A that this is not the case. In a nutshell,

Dyachenko and Allenby (2020) classify observations based on relative fit compared to *other observations*.

Finally, while the biasing role of measurement error in the context of inference about what may be a causal indirect effect has long been recognized (Baron and Kenny, 1986; Cole and Maxwell, 2003), along with latent variable modeling as a viable solution (Iacobucci et al., 2007; Iacobucci, 2008; Cole and Preacher, 2014; Muthén and Asparouhov, 2015) , we need to overcome measurement as an impediment to model discovery.

MEASURING EVIDENCE FOR MEDIATION

This section lays out the individual components of the proposed procedure to measure evidence for mediation in detail. We first summarize the basic causal identifiability argument we rely on, and briefly revisit why this argument has been dismissed as impractical in the past. We then discuss the role of measurement error in the process and introduce latent variable modeling. Finally, we introduce Bayes factors as a means to operationalize the causal identifiability argument in this context, including a numerically reliable computational approach.

A basic result in the literature on model discovery establishes the unique identification of the causal chain from X through M to Y , i.e., full causal mediation from data. In a nutshell, the possibility to discover the presence of such a causal chain from data follows from the insight that if the only connection from X to Y is through M , conditioning on M will leave X and Y independent. The only other models that generate the same conditional independence relationships are a model (i), in which Y causes M which in turn causes X , i.e., the causal chain with reversed causal direction, and a model (ii), where M acts as a common cause to both X and Y . However, random assignment of X rules out both of these models. In addition, a model without a causal effect from X to Y , i.e., unconditional independence between X and Y may also yield conditional independence. However, in connection with evidence for a (total) causal effect from X to Y , conditional independence between X and Y given M uniquely establishes the causal chain (Glymour, 2001, p. 33). In other words, evidence for a (total) causal effect from X to Y strengthens the if-statement from above to: if-and-only-if the only connection from X to Y is through

M, conditioning on M will leave X and Y independent.⁴ From now on, when we mention the unique mapping from conditional independence between X and Y given M onto the causal chain from X through M to Y, we always implicitly assume randomly assigned X and a statistically reliable total effect from X to Y.

Case	Data generating model	Significant indirect effect	Conditional independence
1	Full mediation	✓	✓
2	Partial mediation	✓	x
3	M-Y unobserved confounder	✓	x
4	M-Y reverse causal direction	✓	x

Table 2: Ambiguity of the significant indirect effect and unique model identification from conditional independence.

In the context of the earlier illustrative simulations, conditional independence between X and Y given M thus rules out all cases but the causal chain in Figure 2 (see Table 2). Relatedly, conditional independence rules out cases where variation in putative M that is orthogonal to X (i.e., variation in U_M) is not causally related to Y. For example when putative M is just another indicator for a latent Y, or when putative M is just a reflection of X, i.e., a mere manipulation check.

It is noteworthy that [Baron and Kenny \(1986\)](#) already state that evidence for full mediation (i.e., evidence for a causal chain), is the "strongest demonstration of mediation" (see also [MacKinnon, 2012](#), p. 395; [Shrout and Bolger, 2002](#)). Accordingly, many behavioral researchers agree that full mediation is the gold standard (e.g., [Zhao et al., 2010](#); [MacKinnon, 2012](#)). However, in practice, published mediation studies often fail to even report estimates of the conditional direct effect.

Different arguments lead to de-emphasizing, if not ignoring, the search for conditional independence in mediation models. The first argument relates to the drawbacks of null hypothesis significance testing (NHST). Specifically, failing to reject the absence of a direct effect using NHST (failing to reject $H_0: \beta_3 = 0$ in Figure 1) is consistent with full mediation. However, by construction, NHST does not provide a measure of data based evidence in support of full mediation because the probability of a p-value is equal to its

⁴The event that different non-zero effects, associations cancel to *exactly zero* mimicking conditional independence has probability measure zero ([Pearl, 2009](#), p.150-151). Put differently, consistent evidence for conditional independence is never spurious.

value under the null hypothesis. For this reason [Preacher and Kelley \(2011\)](#) criticize full mediation as "not expressed in a meaningfully scaled metric".

Second, statistical tests of direct effects often suffer from low power, high sampling variability ([MacKinnon et al., 2000](#); [Shrout and Bolger, 2002](#); [Fritz and MacKinnon, 2007](#); [Hayes, 2009](#); [Rucker et al., 2011](#); [Kenny and Judd, 2014](#)). Therefore the absence of a direct effect as measured by these tests could be due to a lack of information in the data. Third, pointing to measurement concerns, some authors advise against testing the direct effect altogether. Specifically, [Rucker et al. \(2011\)](#) argue that "The impossibility of perfect measurement, in and of itself, suggests that one cannot ever claim to have established complete mediation".

Classical additive measurement error in the mediator M complicates standard mediation analysis following [Baron and Kenny \(1986\)](#) or [Preacher and Hayes \(2004\)](#) in two ways: first, it biases the indirect effect estimates, specifically $\hat{\beta}_2$, towards zero because it inflates the mediator's variance. Second, the combination of unbiased $\hat{\beta}_1$ with biased $\hat{\beta}_2$ necessitates bias in the direct effect estimate $\hat{\beta}_3$ —away from zero when there is no causal direct effect β_3 ([Kahneman, 1965](#)). The latter confounds conditional independence and obfuscates full mediation at the causal-theory level. Similar biases arise from using a categorized version of a continuous mediator, such as measures from a rating scale in the analysis (e.g., [Iacobucci, 2008](#)).

Conditioning Variable	M	m^*	$\tilde{m}$
$\hat{\beta}_3$	-0.04 (0.04)	2.45*** (0.08)	0.67*** (0.07)

Table 3: Direct effect estimates when conditioning on the mediator M , on M afflicted by measurement error, m^* , and discretized M , $\tilde{m}$ (standard errors)

We illustrate these biases with a numerical example. We simulate 500 observations from the mediation model shown in Figure 2, where conditional independence holds. Next we simulate noisy observations of M by adding measurement error $U_{m^*} \sim N(0, \sigma_{m^*}^2)$, where $\sigma_{m^*}^2 = 5$, to M . We use the symbol m^* for the noisy observations. The second column in Table 3 summarizes how conditioning on m^* instead of M biases the estimate of the direct effect β_3 away from zero, leading to a spurious rejection of the null hypothesis $\beta_3 = 0$ and thus of full mediation. Finally, we discretize M to observations $\tilde{m}$ on a 7-point rating scale based on cut-off values $\{-2, -0.5, 0.5, 2, 4, 5\}$. Conditioning on $\tilde{m}$ instead

of M again results in a reliably positive estimate of β_3 (see the last column of Table 3), obfuscating full mediation as the underlying causal model. We next introduce explicit models of measurement error to avoid spurious evidence against a causal chain.

Measurement error models

We use latent variable models (LVMs) to accommodate both classical measurement error as well as categorized measures of continuous mediators. Different from previous research, LVMs here serve as a building block to model discovery. Specifically, we need LVMs to avoid spurious evidence against the causal chain from X through M to Y as illustrated above. Empirically, we take advantage of the fact that the vast majority of applications use multiple indicator variables for putative M and Y.

Classical measurement error We require at least two indicators (i.e., two observed measures) for M to empirically identify measurement error independently from a potential direct effect from X to Y (β_3). With one indicator, the model is only set-identified (see online Appendix B for a detailed characterization of the identification problem). The intuition is that the (measurement) error variances of multiple M-indicators and the residual variance in latent M, after conditioning on X, are jointly identified whereas a single M-indicator cannot distinguish between residual variance in M and measurement error variance in the indicator.⁵ Regarding Y, multiple indicators are optional for (point) identification. However, when the mediator is indirectly observed through imperfect measures and there only is one indicator for Y, the conceptual distinction between M and Y no longer directly follows from the model. This is because the Y-indicator can be equivalently viewed as "just another" indicator for M or as a measure of Y. The degree to which this is a problem depends on the prior conceptual knowledge about M and Y as variables.

We model classical measurement error in the usual way where the structural equation,

$$M = \beta_{0M} + \beta_1 X + U_M, \quad (4)$$

⁵Different from the "three indicator rule" for identification of latent variable models we only need two indicators once we condition on X as a cause of latent M (Bollen, 1989), which we can always do in mediation analysis. In other words, the latent mediator version of regressing M on X (Baron and Kenny, 1986) is identified already with only two indicators, and before bringing Y to bear on identification.

defines the prior for latent M and the structural equation,

$$Y = \beta_{0Y} + \beta_2 M + \beta_3 X + U_Y, \quad (5)$$

coupled with the measurement model in Equation 6 define the likelihood for latent M.

$$\begin{aligned} m_1^* &= M + U_{m_1^*} \\ m_2^* &= \lambda_{20} + \lambda_{21} M + U_{m_2^*} \\ &\dots \\ m_K^* &= \lambda_{K0} + \lambda_{K1} M + U_{m_K^*} \end{aligned} \quad (6)$$

The intercept and the slope in the measurement equation for the first indicator m_1^* are fixed to zero and one for identification. The definition of classical measurement error implies that error terms $U_{m_1^*}, U_{m_2^*}, \dots, U_{m_K^*}$ are independently, but not necessarily identically, normally distributed with mean zero. The intercepts $\lambda_{20}, \dots, \lambda_{K0}$ capture bias in indicators relative to the first indicator and the slopes $\lambda_{21}, \dots, \lambda_{K1}$ differential scaling.

Similarly, structural Equation 5 defines the prior and the measurement model in Equation 7 the likelihood for latent Y.

$$\begin{aligned} y_1^* &= Y + U_{y_1^*} \\ y_2^* &= \tau_{20} + \tau_{21} Y + U_{y_2^*} \\ &\dots \\ y_L^* &= \tau_{L0} + \tau_{L1} Y + U_{y_L^*} \end{aligned} \quad (7)$$

Again, $y_1^*, \dots, y_L^*$ are observed indicators and definitions, explanations regarding error terms, intercepts, and slopes from above apply. Figure 5 summarizes the model in the form of a DAG.

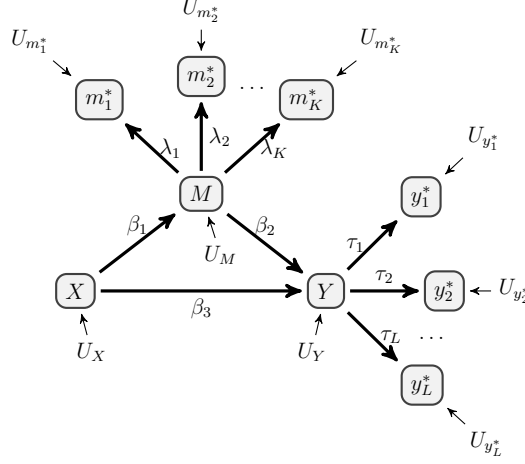


Figure 5: Mediation model with multiple indicators for M and Y

Discretized measures When observed indicators are discretized reflections of M (and Y), such as e.g., measures on rating scales, we add an additional layer to the model in Figure 5. This layer models how unobserved continuous indicators (m_i^* and y_i^*) are discretized into observed indicators ($\tilde{m}_i$ and $\tilde{y}_i$). In the context of rating scales we model the link between discrete observations and unobserved continuous indicators as ordered probit. Using the extended measurement model for the latent mediator M as an example, Equation 6 changes to

$$\begin{aligned}
 \tilde{m}_1 &= \text{OrdProbit}(m_1^*, C_{m_1}), & m_1^* &= M + U_{m_1^*} \\
 \tilde{m}_2 &= \text{OrdProbit}(m_2^*, C_{m_2}), & m_2^* &= \lambda_{20} + M + U_{m_2^*} \\
 &\dots & & \\
 \tilde{m}_K &= \text{OrdProbit}(m_K^*, C_{m_K}), & m_K^* &= \lambda_{K0} + M + U_{m_K^*}
 \end{aligned} \tag{8}$$

Here $C_{m_1}, C_{m_2}, \dots, C_{m_K}$ are indicator-specific ordered cut-points that accommodate a potentially diverse collection of discrete indicators. Note that indicator specific cut-points accommodate heterogeneous (across indicators) non-linear relationships between latent M and the collection of observed indicators and, as a special case, indicator specific scaling. In this way, we capture *all* non-linear relationships between X, the collection of M-indicators, and the collection of Y-indicators that keep with the prior ordering of observed indicator response categories and the uni-dimensional dependence structure between M-indicators (Y-indicators).

Without loss of generality, we constrain all slope parameters $\lambda_{11}, \dots, \lambda_{K1}$ connecting M to indicators (cf. Equation 6) as well as the unique variance of M , σ_M^2 to one for identification. We identify intercepts $\lambda_{20}, \dots, \lambda_{K0}$ by constraining one cut-point for each indicator to zero. As a consequence, the first indicator m_1^* again defines the intercept of latent M . The identification of the generalized-linear measurement model for latent Y proceeds analogously.

Finally, we note that if the measurement problem is known to result from observing the discretization of a perfect indicator, i.e., of M itself, the model is already identified with one indicator only. However, it seems that perfect measurement (up to discretization) is bound to be the rare exception empirically. We do not develop models for inherently discrete mediators in this paper for a lack of application. The general framework, however, applies to inherently discrete mediators as well.

We discuss the details of Bayesian inference for these models using Markov-Chain-Monte-Carlo (MCMC) in Appendix G. There, we also provide a comprehensive overview of the default subjective priors we employ (see Equation 19). All subjective priors except that for the structural coefficient β_3 are standard proper but diffuse priors. We discuss the need for a different prior for β_3 in the context of model discovery in detail in the next section.

Measuring evidence for mediation using Bayes factors

We previously discussed the uniqueness of the mapping from conditional independence between X and Y given M onto the causal chain in which X transmits its effect onto Y only through M . In the context of the LVMs introduced in the previous section, conditional independence corresponds to $\beta_3 = 0$.⁶ A useful measure of potential data based *support* for $\beta_3 = 0$ should not only fail to reject this hypothesis if it is true, but deliver consistently increasing evidence for $\beta_3 = 0$ as more data information becomes available. As we develop and demonstrate next, the Bayes factor comparing a model M_0 constrained to $\beta_3 = 0$ to a model M_1 with unconstrained β_3 fulfills this requirement.

In general, Bayes factors can symmetrically measure data-based evidence both in support and against a null hypothesis. This contrasts usefully with NHST that can

⁶More precisely, $\beta_3 = 0$ only expresses generalized conditional mean independence. However, while it is theoretically possible to have dependence even under generalized conditional mean independence, this case seems practically irrelevant.

only fail to reject a null hypothesis. Moreover, the probability of falsely rejecting the null hypothesis, i.e., the probability of a type I error is constantly equal to the chosen significance level, independent of the amount of data information by construction of NHST. Because of this advantage, Bayes factors have recently gained popularity in applied work (e.g., Rouder and Lu, 2005; van Ravenzwaaij et al., 2011; Annis et al., 2015; Matzke et al., 2015).

Bayes factors are defined as ratios of so called marginal data densities and measure the relative marginal density of the same data (y) given two different models M_0 and M_1 (Equation 9).

$$BF_{01} = \frac{p(y|M_0)}{p(y|M_1)} \quad (9)$$

Computing the marginal data densities in the numerator and the denominator of Equation 9 often is a very difficult problem. In this context, we contribute a numerically stable computational approach that we detail in Online Appendix C. This appendix also illustrates numerical stability and efficiency.

Our computational approach first leverages the fact that in our application the target ratio of marginal densities pertains to two nested model and thus can build on the Savage-Dickey ratio. The Savage-Dickey ratio re-expresses the ratio in Equation 9 as the ratio of the normalized *marginal* posterior density of β_3 in the unrestricted model M_1 over the subjective prior density of β_3 , both evaluated at zero. Second, we leverage a *conditionally* conjugate prior distribution for β_3 that gives rise to a full conditional posterior distribution of known form. In the context of the LVMs introduced in the previous section this becomes possible by augmenting latent variables as part of our MCMC (e.g., Tanner and Wong, 1987).

Finally, we obtain reliable evaluations of the required *marginal* posterior density of β_3 at the value zero by Monte Carlo integration over the joint posterior of all parameters other than β_3 . We achieve this at negligible additional computational costs and high numerical accuracy by saving the full conditional moments indexing individual draws of β_3 when sampling from the posterior of the unrestricted model M_1 . The Monte Carlo integration yielding the marginal posterior density then simply recycles the saved full conditional moments.

The magnitude of the Bayes factor is a measure of evidence. Kass and Raftery (1995) provide guidelines for how to interpret Bayes factors in terms of the amount of evidence

supporting the model in the numerator of Equation 9 (see Table 4). By symmetry of the Bayes factor, a value smaller than e.g., $1/100$ would be decisive evidence *against* the model in the numerator.

BF_{01}	Evidence in favor of H_0
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

Table 4: Interpreting Bayes factor values

However, as we illustrated previously, violations of conditional independence can have multiple indistinguishable reasons, such that evidence supporting M_1 ($\beta_3 \neq 0$) is inconclusive in the absence of prior knowledge about the data generating causal structure.⁷ In contrast, evidence supporting M_0 , in combination with a credible total effect ($X \rightarrow Y$), reveals the data generating causal model, i.e., (full) mediation, when X is randomly assigned. Next, we use an illustrative simulation to show how the Bayes factor can consistently measure evidence for the presence of a causal chain and hence for causal mediation. The simulation also shows how measurement error that is unaccounted for obscures the presence of a causal chain, i.e., the implied conditional independence relationship.

Illustrative simulation We illustrate with simulated data from the model in Figure 5 with 2 indicator variables for M and Y each, and the following data generating parameter values: $X \sim U(0, 1)$, $\beta_1 = 1$, $\beta_2 = 3$, $\beta_3 = 0$, $\{\lambda_{10}, \lambda_{11}\} = \{\tau_{10}, \tau_{11}\} = \{0, 1\}$, $\{\lambda_{20}, \lambda_{21}\} = \{1.5, 1\}$, $\{\tau_{20}, \tau_{21}\} = \{1.5, 1\}$, $Var(U_M) = 1$, $Var(U_Y) = 1$, $Var(U_{m_1^*}) = .3$, $Var(U_{m_2^*}) = .6$, $Var(U_{y_1^*}) = .6$, $Var(U_{y_2^*}) = .3$. We simulated and analyzed 1000 data sets each for sample sizes $N = 50, 200$, and 1000 . Note that because $\beta_3 = 0$ we have a causal chain from X through M to Y that may however be obfuscated by the measurement error in indicators (m_1^*, m_2^*) of M . The purpose of the simulation is to show how Bayes factors measure evidence supporting full mediation once we account for measurement error in the mediator, and to showcase consistency of the Bayes factor. The simulation also serves as a simple example of a power analysis for model choice.

⁷Correct inference using the approaches proposed by [Imai et al. \(2010\)](#) and [Zhang et al. \(2009\)](#) at least requires definite knowledge about the causal order of measured variables, i.e., putative M and Y .

Figure 6 summarizes the distribution of Bayes factors comparing the constrained model ($\beta_3 = 0$) to its unconstrained counter-part across simulated data sets. In the left panel we report results from models that do *not* account for measurement error in the mediator and instead estimate the simple model using averages of the indicators $\{m_1^*, m_2^*\}$ and $\{y_1^*, y_2^*\}$ as composite measures of latent M and Y, respectively.

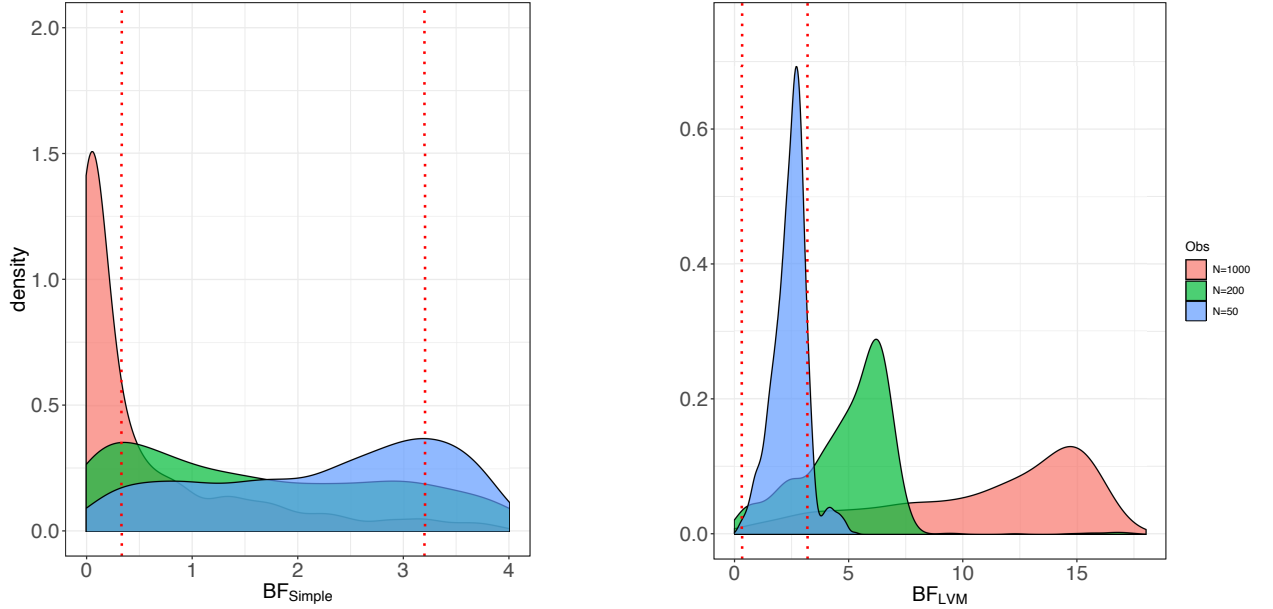


Figure 6: Bayes factor distribution conditional on different sample sizes in a full mediation model before (left panel) and after (right panel) controlling for measurement error. Red dotted vertical lines bound intervals of inconclusive Bayes factor values.

The right panel summarizes results from models that do account for measurement error in the mediator, i.e., assess conditional independence between latent M and Y. In each panel there are three distributions of Bayes factors (across simulated data sets), one each for the three sample sizes investigated. The red dotted vertical lines in the figures demarcate the lower and upper end of the interval in which Bayes factors do not support a decision between models (see Table 4).

When we do not account for measurement error, Bayes factors for $N=50$ and $N=200$ are spread out in the interval $[0, 3.2]$ with $N=200$ having more mass in $[0, 0.32]$, which is the region with at least substantial evidence *against* the constraint $\beta_3 = 0$ (see the left panel of Figure 6).⁸ Larger data sets ($N=1000$), however, yield Bayes factors concentrated

⁸Note that border values 0.32 and 3.2 in Kass and Raftery’s guidelines should not be treated as determinate thresholds for rejecting/accepting conditional independence. In fact the wide support of Bayes factor distributions in Figure 6 points to the inherent uncertainty in limited data. Therefore, we encourage collecting more data when $0.32 < BF < 3.2$, but do not advocate overconfident claims based on Bayes factors marginally higher than 3.2 or lower than 0.32 (see Amrhein et al., 2019).

at very small values that translate into decisive evidence *against* $\beta_3 = 0$. Thus, we see how measurement error can obfuscate the causal structure that generated the data.

Once we control for measurement error using the latent variable model introduced in Section 4, Bayes factors for $N=200$ and 1000 , yield substantial evidence supporting the constraint $\beta_3 = 0$, i.e., unequivocal support for the causal chain from X through (latent) M to (latent) Y (full causal mediation). It is noteworthy that the amount of evidence supporting the constraint *increases* in the sample size, when the constraint holds meeting our consistency requirement.⁹ This is a demonstration of how Bayes factors measure the amount of *data based* evidence *supporting* a constraint. When the data are less informative because of a small sample size ($N=50$), Bayes factors do not support a firm decision between models, as they should given limited evidence.

Finally, the illustrative simulation reflects [Rucker et al. \(2011\)](#)’s observation that one is likely to spuriously *reject* conditional independence in the presence of measurement error. It also shows that testing conditional independence between latent M and Y is a viable strategy to overcome the complications from potential measurement error. Moreover, it illustrates how Bayes factors work as the sought after metric of the degree of data based support in favor or against conditional independence given a model specification (cf. [Preacher and Kelley, 2011](#)). In contrast, NHST fails to provide a measure of support for the null hypothesis by construction and thus cannot per se distinguish between an inconclusive result for a lack of power for detecting $\beta_3 \neq 0$ and data based evidence in favor of $\beta_3 = 0$.

The role of the subjective prior distribution Bayesian inference is defined as the process of updating prior knowledge with data. When subjective prior need to be specified for parameters *in a given model*, a standard recommendation is to use weakly informative, diffuse but proper priors in the absence of specific subjective prior knowledge or theory to guide their formulation. However, our goal is not only to update prior knowledge about parameters in a given model but to *learn about the underlying model structure*. The key parameter in this context is β_3 where increasing data-based support for $\beta_3 = 0$ (i.e., generalized mean independence between X and Y given M) unequivocally identifies the causal chain from X through M to Y as the underlying causal model (full causal

⁹In Online Appendix C we show that the Deviance Information Criterion (DIC) as an alternative measure for model comparison ([Gelman et al., 2013](#)) fails this consistency requirement.

mediation). As we explain next, non-trivial data-based evidence for $\beta_3 = 0$ requires a different default prior specification.¹⁰

It is important to realize that a diffuse subjective prior for β_3 not only expresses a lack of prior knowledge about β_3 , but also the firm belief that β_3 could be very large (in an absolute sense). Recalling that our target Bayes factor can be re-expressed as the ratio of the marginal posterior of β_3 over its subjective prior, both evaluated at zero, any concentration of posterior mass closer to zero will result in a Bayes factor supporting $\beta_3 = 0$ under a diffuse subjective prior. Put differently, a standard weakly informative diffuse prior for β_3 is only weakly informative *given a fixed model structure*, but highly informative in the context of model choice. It follows that for non-trivial model choice, we need to avoid diffuse priors for β_3 .

In contrast, specifying a direct effect prior that is concentrated at zero expresses the belief that a possible direct effect β_3 is close to zero. With such a prior, additional data based information that the effect is in fact zero needs to be strong for the Bayes factor to result in substantial evidence for $\beta_3 = 0$. The prior structural belief that a causal chain may be at work naturally translates into a direct effect prior concentrated at zero, i.e., the disbelief in arbitrarily large direct effects. And it is such a more concentrated prior that translates into non-trivial data-based evidence either supporting or refuting conditional independence.¹¹

In our simulations and in our re-analyses of published studies presented below, we use a normal prior for β_3 centered at zero and with prior variance equal to 1 ($\beta_3 \sim N(0, 1)$). This setting has been advocated as a default or reference prior in psychology (see e.g., [Rouder et al., 2012](#); [Wagenmakers et al., 2011](#); [Rouder et al., 2009](#)). This prior expresses that direct effects absolutely larger than two are outlying events, i.e., have a prior probability of less than five percent.

To further probe into the reasonableness of this subjective prior setting, we surveyed recently published mediation studies in *Journal of Marketing Research* (JMR), *Journal*

¹⁰All other parameters rely on standard weakly informative, diffuse but proper priors as detailed in Equation 19 in Appendix G.

¹¹The sensitivity of Bayes factors to the choice of subjective prior is often viewed as a drawback. However, it is precisely this property that guarantees the consistency of Bayes factors in our application. It is easy to show that for any non-degenerate, continuous and proper subjective prior for β_3 (with positive support at $\beta_3 = 0$) the Bayes factor BF_{01} investigated here converges to infinity when in fact $\beta_3 = 0$, and to zero when $\beta_3 \neq 0$, in the limit of infinite data (e.g., [O’Hagan and Forster, 1994](#), p. 196). This property follows from the convergence of the posterior to a point mass at $\beta_3 = 0$ (at some value $\beta_3 \neq 0$) in the limit of infinite data.

of *Consumer Research* (JCR), *Journal of Personality and Social Psychology* (JPSP), and *Psychological Science* (PS) that report estimates of β_3 .¹² Figure 7 summarizes the distribution of the 114 point estimates we found.

In line with the advocated default prior, we see that estimates concentrate around zero with a few outlying values. The variance across our 114 data points is 5.01 and thus larger than the default prior variance. While this may lead to shrinking some of our estimates of β_3 towards zero in the unconstrained model M_1 , it ensures non-trivial evidence measures supporting or refuting conditional independence from our target Bayes factor. Further details on the choice of prior for β_3 , including its assessment in the context of LVMs are reported in Appendix D.

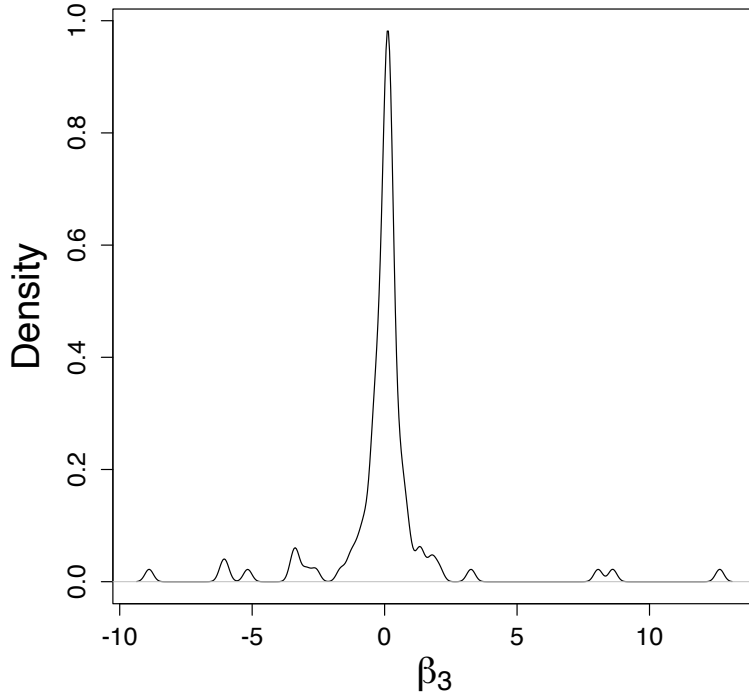


Figure 7: Distribution of 114 direct effect values from recently published studies in JMR, JCR, JPSP, and PS.

Comprehensive simulation study To examine the robustness of the suggested procedure that combines latent variable modeling, Bayesian inference, and Bayesian model choice under different conditions that might arise in applied research, we conduct a simu-

¹²Studies published in *Journal of Marketing Research* between February 2014 and December 2018, studies published in *Journal of Consumer Research* and *Journal of Personality and Social Psychology* in 2017, and studies published in *Psychological Science* between January 2015 and September 2018 with open-data badge.

lation study. The simulation study investigates four different data generating structures: i) full mediation (see Figure 2), ii) partial mediator (see Figure 1), iii) structures with an unobserved M-Y confounder and no mediation (see Figure 3), and finally, iv) the case of a misspecified causal ordering of measured variables (see Figure 4).

We investigate small (N=50), medium sized (N=350), and large datasets (N=1000). The medium sample size is roughly in line with the average sample size (N=319) in the pool of published studies we re-analyze in the following section. Indirect and direct effect sizes are varied in a range following prior literature (Tofghi and Kelley, 2020), and consistent with the range of estimated effects in our re-analysis of published studies. We vary the reliability of mediator and outcome indicator variables following Fritz et al. (2016). We include conditions with non-normal error distributions following Finch et al. (1997) and Tofghi and Kelley (2020). Finally, we vary the number of indicator variables mimicking the characteristics of published studies we re-analyze later. In total, our simulation study features 3,024 cells. We base our results on 100 data replications in each cell for a total of 302,400 data sets, MCMC runs and Bayes factor computations.

Model $\Downarrow$ Sample $\Rightarrow$	N50	N350	N1000
Full mediation	0.2%	77.9%	89.4%
Partial mediation	99.9%	95.6%	99.9%
M-Y unobserved confounder	100%	98.8%	100%
M-Y misspecified direction	99.9%	99.9%	100%

Table 5: The percentages of Bayes factor values in the simulation study providing at least substantial evidence for conditional independence (in the full mediation model), not providing evidence for conditional independence (in the alternative models). See Online Appendix E for full details.

Table 5 summarizes the simulation study in terms of Bayes factor accuracy across different data generating structures and sample sizes, marginalizing over all other conditions in our simulation. The percentages displayed in Table 5 show, for each data generating structure and sample size, how often the procedure at least substantially supports conditional independence where it should (first row), and how often it does not support conditional independence where it should not (all other rows). Confirming our previous discussion and illustration, smaller datasets (N=50) do not contain enough information for Bayes factor to substantially support conditional independence. Reassuringly, we do not see false positives either (see the column "N50" in Table 5).

In medium sized datasets ($N=350$) we find at least substantial evidence for causal mediation in roughly 80% of datasets generated from an unconfounded full mediation structure. The sensitivity increases to about 90% in large samples ($N=1000$). These results suggest $N=350$ as a reasonable lower bound for studies aiming at potentially substantiating full causal mediation.¹³ Finally, when conditional independence is violated in the data generating structure, Bayes factors reliably fail to support it, i.e., do not produce false positives.

Table 6 provides a more detailed summary of results. The table presents coefficients from regressing the percentage of Bayes factors providing at least substantial evidence supporting (column "Full"), or refuting (all other columns) conditional independence on experimental factors of the simulation study. The columns correspond to the same four data generating structures as described previously. The strength of indirect effect manipulation corresponds to the strength of the unobserved background connection between M and Y in column 3 ("M-Y confounder"), and to the strength of the actual causal connection in column 4 ("M-Y direction"), where the estimated model confuses the causal ordering of measured variables.

The main result from our comprehensive simulation is that sample size (N) is the key to measuring evidence supporting or refuting conditional independence. Other factors are either insignificant, despite 100 replications per simulation cell, or smaller in order of magnitude. More dependence between measured variables increases the chance of substantial evidence against conditional independence in columns 3 and 4. The smaller negative effect of the indirect effect strength on Bayes factor accuracy in full and partial mediation structures (columns 1 and 2) is due to the induced correlation between X and M that translates into less statistical information about the direct effect. Larger measurement error variance (σ_{m*}, σ_{y*}) in general reduces Bayes factor accuracy.¹⁴ More Y and M indicators (M_{ind}, Y_{ind}) in general lead to higher accuracy in Bayes factors. Finally, non-normality of the error terms does not have a discernable effect on the assessment of conditional independence using our Bayes factor measure. Further details regarding our simulation study are available in Online Appendix E.

¹³In addition to this guideline, we provide a Bayesian power analysis function in the R package accompanying this paper. The function simulates Bayes factor distributions under the hypothesis of full, unconfounded causal mediation for different parameter settings and sample sizes.

¹⁴The small positive effect of σ_{y*} in the full mediation condition relates to an interaction with sample size. After controlling for this interaction, the negative main effect of σ_{y*} is recovered.

Table 6: Results of the regression, where the dependent variable is the percentage of studies in the simulation study where the procedure correctly identifies (rejects) conditional independence in datasets generated from the full mediation model (alternative models). The (dichotomized) regressors are as follows: $\beta_1\beta_2$ = indirect effect (baseline = small effect), σ_{m*} = mediator measurement error variance (baseline = low variance, 1 = high variance), σ_{y*} = outcome measurement error variance (baseline = low variance, 1 = high variance), N = sample size (baseline N = 50), M_{ind_4} = 4 mediator multiple indicators (baseline = 2), Y_{ind_4} = 4 outcome multiple indicators (baseline = 2), Non-normal = measurement errors with non-normal distributions (baseline = normal, 1 = moderately non-normal, 2 = severely non-normal)

	<i>Dep.var.: % substantial evidence for/against cond. indep.</i>			
	Full	Partial	M-Y confounder	M-Y direction
	(1)	(2)	(3)	(4)
$\beta_1\beta_2 - Medium$	-1.042** (0.437)	-1.847 (1.127)	2.877** (1.189)	2.285*** (0.583)
$\beta_1\beta_2 - Large$	-2.132*** (0.437)	-3.493*** (1.127)	5.516*** (1.189)	2.688*** (0.583)
σ_{m*}	-0.694* (0.357)	-1.486 (0.920)	-1.650* (0.971)	-3.199*** (0.476)
σ_{y*}	1.472*** (0.357)	-1.745* (0.920)	-7.171*** (0.971)	-0.560 (0.476)
N_{350}	77.653*** (0.437)	55.899*** (1.127)	27.697*** (1.189)	64.674*** (0.583)
N_{1000}	89.174*** (0.437)	75.646*** (1.127)	35.225*** (1.189)	67.424*** (0.583)
M_{ind_4}	1.333*** (0.357)	5.588*** (0.920)	-0.792 (0.971)	1.088** (0.476)
Y_{ind_4}	-0.843** (0.357)	0.236 (0.920)	4.468*** (0.971)	3.116*** (0.476)
$Non - normal_1$	0.486 (0.437)	0.816 (1.127)	0.595 (1.189)	-0.187 (0.583)
$Non - normal_2$	0.278 (0.437)	0.771 (1.127)	0.451 (1.189)	0.639 (0.583)
Constant	0.398 (0.592)	23.028*** (1.526)	63.641*** (1.610)	30.546*** (0.790)
Observations	432	864	1,296	432
R ²	0.992	0.852	0.455	0.976

Note:

*p<0.1; **p<0.05; ***p<0.01

RE-ANALYSIS OF PUBLISHED STUDIES

To showcase the empirical relevance of the proposed methodology, we re-analyze five recently published mediation studies (see online Appendix F for a complete overview of re-analyzed studies published in *Journal of Marketing Research* between February 2014 and December 2018 and *Psychological Science* between January 2015 and September 2018 with open-data badge). In the first two studies, NHST yields results consistent with full mediation. We show how the proposed methodology, and specifically the application of Bayes factors measures scaled evidence supporting full mediation, i.e., provides a metric. The third study we discuss is an example of how measurement error results in biased estimates, obfuscating full mediation. We show how a latent variable model recovers evidence for conditional independence between Y and X given M, strongly confirming full mediation as the data generating causal model.

The fourth study yields results that are only consistent with mediation because of the indeterminate mapping from results consistent with partial mediation to the underlying causal model. Controlling for measurement error does not overcome the indeterminacy in this case. However our analysis illustrates how controlling for measurement error yields an improved estimate of the indirect effect, if someone believes in mediation as the data generating causal model a priori. Finally, in the fifth study, we examine a serial mediation claim using our procedure and arrive at an explanation for the underlying process by ruling out equivalent models. This way, we are able to resolve fundamental uncertainties about the causal ordering of variables made explicit in the originally published analysis.

Empirical Illustration of Bayes Factors as Measures of Evidence Supporting Full Mediation

In the first study we re-analyze here, [Shaddy and Fishbach \(2017\)](#) show that customers ask for more compensation when a product is lost from a bundle than when it is lost in isolation. The reason, the authors argue, is that in the former case, in addition to the loss of the product, the "whole" of the bundle as a gestalt unit is compromised. Hence, the authors propose that the perceived need for replacement of the lost product mediates the effect of its loss from a bundle (vs. in isolation) on the amount of compensation demanded. In an experiment participants were randomly assigned into two groups either imagining

they have purchased three suitcases as a bundle or separately. The participants were then asked how much, in dollars, they think they should be compensated, and how important, on a 7-point rating scale, it would be to replace the missing item. The authors estimate a significant indirect and an insignificant direct effect (see Figure 8a and Table 7).

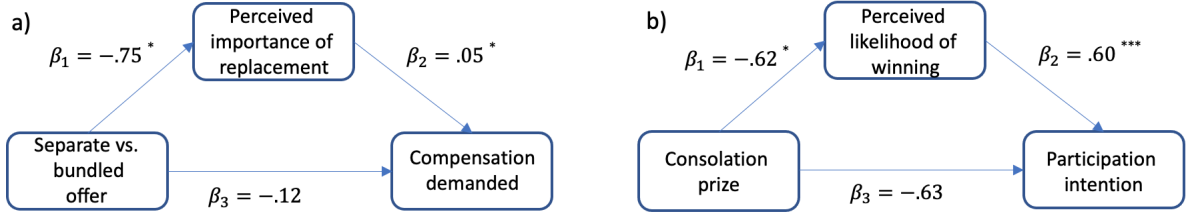


Figure 8: Mediation diagrams from Shaddy and Fishbach (2017) (a), Yan and Muthukrishnan (2014) (b) with standard test results

	Shaddy & Fishbach (2017)	Yan & Muthukrishnan (2014)
NHST	fail to reject full mediation	fail to reject full mediation
Direct Effect Bootstrapped CI	[-0.29 , 0.05]	[-1.33 , 0.09]
Bayes factor	4.9	0.66

Table 7: A comparison of NHST, direct effect bootstrapped confidence interval and Bayes factor in cases where full mediation hypothesis cannot be rejected based on NHST.

While these results are consistent with full mediation, it is hard to tell how much evidence they provide. For example, the asymmetry in the bootstrapped 95% confidence interval may lead to speculations about a lack of statistical information about the direct effect in the data. The Bayes factor we compute for this study points to substantial evidence supporting full mediation (BF=4.9), alleviating this and similar concerns.

Substantively, evidence supporting full mediation, among other competing explanations, rules out a reverse causal direction from demanding a high compensation amount (Y) to reporting a high perceived importance of replacement (M). This is particularly helpful in this study because the perceived importance of replacement was elicited only after asking for the compensation amount. Without the data based insight into the causal ordering from X via perceived importance of replacement to the compensation amount, the time order of measurements could lead to speculations about a different process.

The second study we re-analyze here, by Yan and Muthukrishnan (2014), proposes that consumers' valuations of a lottery-based promotion and intentions to participate in such a promotion are greater when the lottery excludes (vs. includes) consolation prizes. The authors hypothesize that the effect of the consolation prize on participation intention

is mediated by the perceived likelihood of winning. The authors report qualitatively similar estimation results as in the study by [Shaddy and Fishbach \(2017\)](#) discussed above, based on which they claim full mediation (see Figure 8b and Table 7). Even though the NHST results are consistent with full mediation ($95\%CI_{\beta_3} = [-1.33, 0.09]$), the Bayes factor we compute reveals a lack of evidence for full mediation ($BF = .66$). There is just not enough information in the data to rule out alternative models and precisely this lack of information could have caused the insignificant direct effect estimate. As such, this study does not per se deliver data based evidence for mediation. The causal process interpretation of what could be an indirect effect has to rest on subjective prior assumptions about the underlying data generating model structure.

Testing Full Mediation Accounting for Measurement Error

Small Bayes factor values indicate that there is insufficient empirical evidence to pinpoint an underlying model. Very small Bayes factors are evidence against conditional independence and hence *prima facie* evidence against full mediation. However, since measurement error and discretization might confound testing conditional independence between Y and X given M, controlling for these aspects of measurement may still establish full mediation as the causal model generating the data in these instances. Controlling for these two sources of bias is possible when a study features multiple indicators for M and Y. Next, we showcase how this can improve mediation analysis in an empirical example.

[Sussman and O'brien \(2016\)](#) assess whether people take on average a smaller amount of money from the savings they have set aside for their children, education, or retirement (versus vacation, or buying a car), because they perceive it as irresponsible such as e.g., "Taking money from savings that was set aside [for my child] makes me feel like a bad person.". They use bootstrapping to statistically test what could be an indirect effect and find results consistent with mediation ($95\%CI_{\beta_1*\beta_2} = [38.95, 180.83]$). As we discussed in Section 1, this result in isolation is equivocal, i.e., does not identify mediation as the underlying causal model. For example, it seems hard to rule out a priori that respondents answer the questions about the perceived irresponsibility of their withdrawals, at least in part, in reaction to the amount of money they said they would take from a particular savings account. Again, the order of measurements that first elicits the withdrawal amount and then asks about the perceived irresponsibility may induce such a relationship.

Estimating a simple mediation model, the 95% posterior credible interval (PCI) yields a barely insignificant direct effect ($95\%PCI_{\beta_3-Simple} = [-.02, .33]$).¹⁵ A Bayes factor of 2.1 indicates a negligible amount of evidence supporting full mediation from savings account type (X) through perceived responsibility (M) to the withdrawal amount (Y). This leaves room for reverse causation (a $Y \rightarrow M$ effect), and unobserved M-Y confounders.

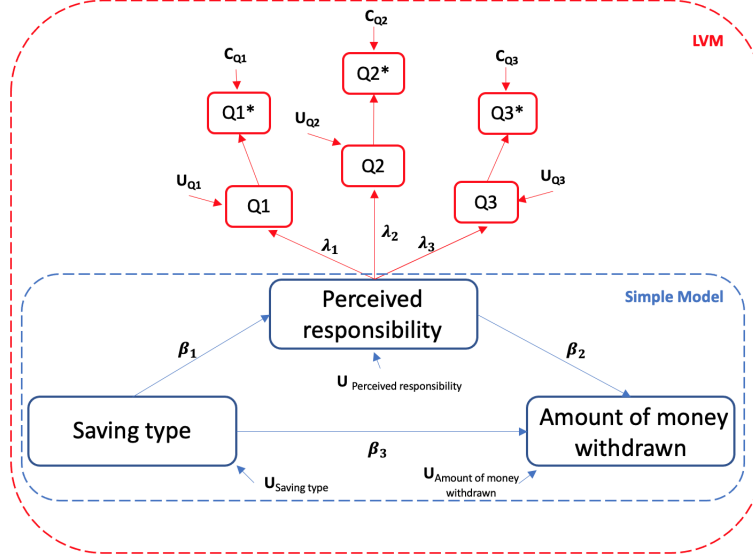


Figure 9: Diagram of Sussman & O'Brien (2016) mediation study. In the simple model, perceived responsibility is a composite measure of the observables and in LVM it is a latent variable.

In search for more accuracy and potentially conclusive results that rule out alternative causal models, we leverage the multiple measures of perceived responsibility in this study. Perceived responsibility is measured using three questions on 7-point rating scales. Using these measures as indicators, we calibrate a latent variable model to probe into the possibility that measurement error and discretization may have affected the results (Figure 9). In contrast, the simple model uses the average of the responses to these questions as a composite measure of the mediator.

Figure 10 illustrates the posterior distribution of the indirect effect through perceived responsibility in the left panel. The right panel shows the posterior distribution of the direct effect. Both panels overlay the respective posteriors from the simple model and from the latent variable model. Comparing posterior distributions between the two models reveals that after accounting for discretization and measurement error, the indirect

¹⁵To keep with our default prior, we standardized the withdrawn dollar amount. In the simple model averages of indicators are taken as error free measures.

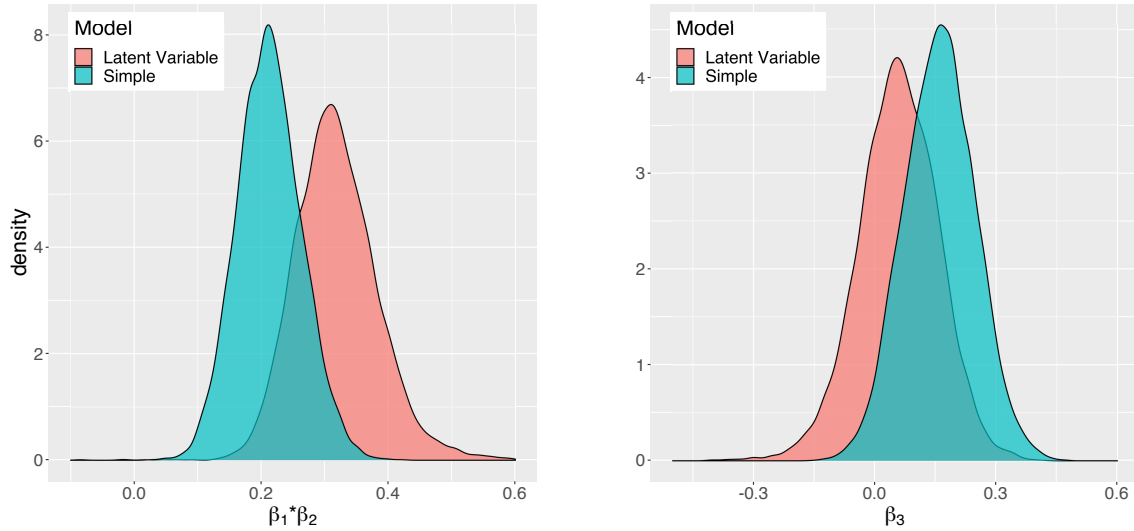


Figure 10: Sussman & O’Brien (2016)’s overlay plots of the indirect effect (left panel) and the direct effect posterior (right panel) before (in aqua) and after (in red) controlling for measurement error and discretization

effect PCI increases from $[.12, .32]$ in the simple model to $[.20, .47]$ in the LVM (see the left panel). In addition, the direct effect estimate, which barely covered zero before accounting for measurement error and discretization, now is closer to zero and its PCI close to symmetrically straddles zero ($95\%PCI_{\beta_3-LVM} = [-0.14, .24]$). In line with these observations, the overall evidence uniquely supporting full mediation as the underlying causal model has increased to substantially supporting this model ($BF = 8.5$). The result is useful because it rules out alternative causal models such as a reverse causal ordering from the withdrawal amount to perceived responsibility or unobserved common causes.¹⁶

Assessing (Potential) Partial Mediation in the Presence of Measurement Error

If a dataset is generated by a partial mediation model, or common unobserved causes link M to Y, Bayes factors will reject conditional independence between Y and X given M, and thus fail to produce evidence supporting full mediation, as they should, and even after controlling for measurement error. An empirical example can be found in [Sela and LeBoeuf \(2017\)](#). These authors hypothesize that people are more likely to buy an upgrade of the product they own if they are not encouraged to compare the upgrade with the

¹⁶If we (ex post mistakenly) reverse the causal ordering of perceived responsibility and withdrawn amount, the Bayes factor drops to $8.7 * 10^{-06}$ which, upon taking the inverse $1/(8.7 * 10^{-06}) = 114044$, is decisive evidence against conditional independence under this alternative causal ordering.

status quo product, and this happens because they tend to perceive the upgrade as more dissimilar to their current version.

Sela and LeBoeuf (2017) conducted an experiment in which they randomly assign participants to the treatment group, where participants were prompted to compare their current mobile phone with an upgraded version of the phone. All participants were then asked to rate the similarity of their phone with the upgrade version, using two questions on 7-point rating scales, and rate how interested they would be to upgrade their phone, again on a 7-point scale. Statistics reported in the paper, based on bootstrapping, suggest a significant indirect effect, based on which the authors concluded mediation is present ($95\%CI_{\beta_1*\beta_2} = [-0.36, -0.06]$).

We use the two (ordinal) indicators of perceived similarity (M) in a latent variable model. Even after controlling for measurement error and discretization of mediator measures, the Bayes factor decisively rejects conditional independence between the experimental condition and the purchase intention after controlling for similarity ($BF = .000006$). However, if there is a strong prior belief that partial mediation (see Figure 1) is the underlying causal model, the 95% PCI of $\beta_1 \cdot \beta_2$ from this model is the preferred estimate of the indirect effect because measurement error and discretization of observed measures are controlled for. In fact, the estimate is more reliably negative than that from the simple model ($95\%PCI_{\beta_1*\beta_2-LVM} = [-.42, -.08]$).

This study presents a case where evidence from the data is insufficient to identify the data generating causal model. However, accounting for sources of bias, namely measurement error and discretization, can yield cleaner, more reliable parameter estimates in a partial mediation model which we may believe to be true a priori.

Testing Serial Mediation

In the last exemplary study we re-analyze, Tonietto and Malkoc (2016) use NHST to obtain results consistent with serial mediation. The results suggest that scheduling a leisure activity (X), makes it feel less free-flowing (M1), which in turn makes that activity feel more like work (M2), which finally reduces subjects' anticipation utility (Y) (see Figure 11a). However, the authors note that such a mediation model cannot in fact establish a causal ordering of variables and mention the possibility that anticipation utility

could be an antecedent to work construal, or that anticipation utility and work construal could be joint dependent variables rather than sequential outcomes of scheduling.

We next show how our suggested approach applies to testing serial mediation and thereby helps rule out some of the alternative interpretations mentioned by [Tonietto and Malkoc \(2016\)](#). For an empirically validated causal interpretation of conditional effects connecting measured variables M1 to M2 and M2 to Y as causes and consequences, we need (i) conditional independence between X and M2 given M1 ($X \rightarrow M1 \rightarrow M2$), (ii) conditional independence between X and Y given M1 ($X \rightarrow M1 \rightarrow Y$), and finally (iii) conditional independence between X and Y given M2 ($X \rightarrow M2 \rightarrow Y$).¹⁷ Conditional independence relationship (i) establishes the indirect effect from X to M2 via M1 as the only causal connection between X and M2, and thus that the effect from M1 to M2 is causal. Relationship (ii) establishes the indirect effect from X to Y via M1 as causal, eliminating the possibility of a direct effect or unobserved non-causal connections between M1 and Y, and relationship (iii) that from X to Y via M2 (again eliminating the possibility of a direct effect and now of unobserved non-causal connections between M2 and Y). Together, these conditional independence relationships uniquely establish the causal chain from X (scheduling a leisure activity) via M1 (a less free-flowing experience) to M2 (work construal), and finally to Y (anticipation utility). Conditional independence between M1 and Y given M2 ($M1 \rightarrow M2 \rightarrow Y$) is implied.¹⁸

The Bayes factors comparing the respective models assuming conditional independence to their dependent alternatives are (i) 2.3, (ii) 12.1, and (iii) 12.9. Hence, the results strongly support the causal models $X \rightarrow M1 \rightarrow Y$ and $X \rightarrow M2 \rightarrow Y$, but fail to support $X \rightarrow M1 \rightarrow M2$. In addition, testing conditional independence between M1 and Y given M2 ($M1 \rightarrow M2 \rightarrow Y$) yields a Bayes factor of only .02, strongly rejecting conditional independence, and hence the unique identification of the serial causal chain $X \rightarrow M1 \rightarrow M2 \rightarrow Y$ as the data generating model. In fact, our results rule out all causal models that feature X, M1, M2, and Y as four conceptually distinct variables. Specifically, the strong support for conditional independence between X and Y given either M1 or M2

¹⁷Recall that all effects from X onto other variables are causal by random assignment of X in the experiment.

¹⁸Relationship (i) establishes a uniquely causal relationship from M1 to M2, relationship (ii) the absence of unobservables connecting M1 and Y, and relationship (iii) the absence of unobservables connecting M2 and Y.

coupled with the failure to support additional independence relationships is at odds with any causal ordering of these *four* variables.

However, from the strongly supported conditional independence implied by $X \rightarrow M1 \rightarrow Y$ and by $X \rightarrow M2 \rightarrow Y$ we learn that the triples $\{X, M1, Y\}$ and $\{X, M2, Y\}$ each *individually* consist of conceptually different variables. The possibility of a causal ordering only vanishes when we consider M1 and M2 as conceptually different. Thus, we investigated a model that uses the indicators of M1 and M2 as measures of one latent mediator M (see Figure 11b), and obtain again strong support for conditional independence as implied by the model $X \rightarrow M(M1, M2) \rightarrow Y$ with a Bayes factor of 17.9.¹⁹

Compared to [Tonietto and Malkoc \(2016\)](#)’s original conclusion, we strongly establish anticipation utility (Y) as a conceptually distinct final outcome. We also conclusively establish mediation by M1 (M2) as the causal transmission mechanism from scheduling a leisure activity or not (X) onto anticipation utility (Y). However, at the same time our results strongly refute the possibility of a causal ordering among M1 and M2, and hence serial mediation. Together with conditional independence between X and Y given either M1 or M2, it follows that M1 and M2 are *not* conceptually distinct in their relation to X and Y.

¹⁹Using the composite measures of M1 and M2 in the model, as it is done in [Tonietto and Malkoc \(2016\)](#) yields qualitatively similar results although with a smaller Bayes factor (BF = 13.76), which is due to unaccounted measurement error and discretization.

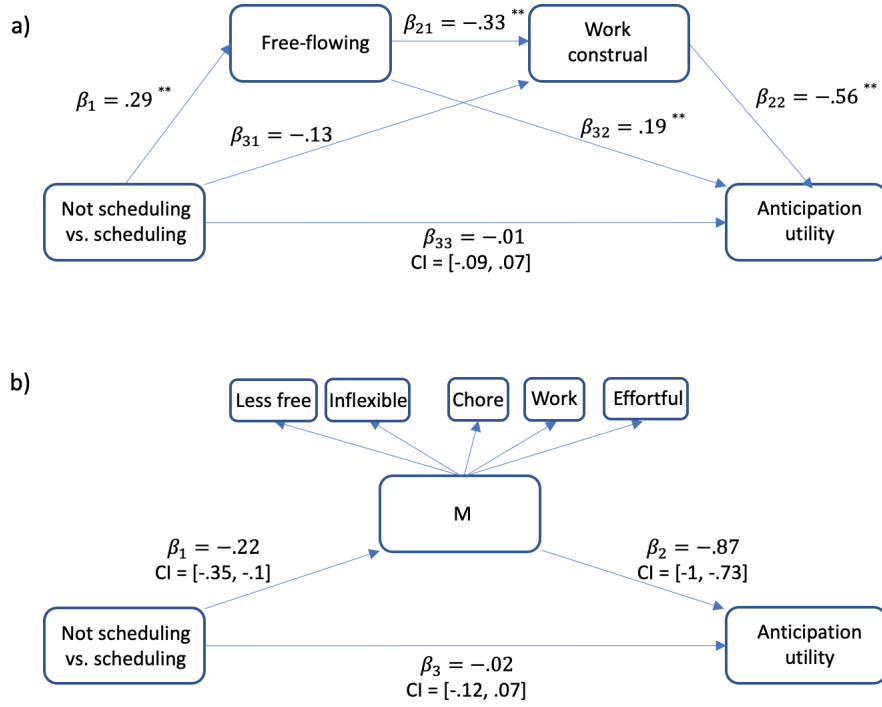


Figure 11: Mediation diagrams of [Tonietto and Malkoc \(2016\)](#) assuming a serial mediation model (a), and a mediation model with a latent mediator (b).

The proposed procedure combines Bayes factors with latent variable models to identify causal mediation in the data by examining conditional independence. The above results point to the importance of each of the procedure's components as well as the procedure as a whole. The data from [Yan and Muthukrishnan \(2014\)](#) and [Shaddy and Fishbach \(2017\)](#) showcase empirically how the suggested Bayes factor provides a metric of support for full causal mediation, i.e., distinguishes between an insignificant direct effect resulting from a lack of power or sample variability as in [Yan and Muthukrishnan \(2014\)](#) and strong data-based support for conditional independence as in [Shaddy and Fishbach \(2017\)](#). The second application from [Sussman and O'brien \(2016\)](#) provides an empirical example for obfuscation from measurement error, and showcases our ability to measure support for conditional independence between X and Y given latent M.

The third application from [Sela and LeBoeuf \(2017\)](#) exemplifies cases where our procedure results in evidence against conditional independence. In such cases, as previously summarized in Table 2, ambiguity about the underlying causal model cannot be resolved. However if prior knowledge rules out unobserved confounders as well as a reversed $Y \rightarrow M$ causal ordering, the estimated indirect effect can be interpreted as causal. In addition,

based on a definite prior causal ordering of measured variables, sensitivity analyses as suggested in [Imai et al. \(2010\)](#) become a possibility.

Finally, the last application shows how the procedure can help resolve fundamental uncertainty about the causal ordering and nature of variables under investigation. We reduce a number of prior possible causal structures suggested in [Tonietto and Malkoc \(2016\)](#) to only one, which derives from the conditional independence relations we find in the data. This way our approach helps researchers elucidate the causal structure underlying their data as compared to focusing on functions of coefficients in a given model of serial mediation(e.g., [Tofghi and Kelley, 2020](#)).

SUMMARY

Figure 12 presents a flowchart of the proposed approach towards mediation analysis. The procedure applies to the setting where X is randomly assigned and M and Y are measured (so called measurement-of-mediation designs). Assuming a statistically reliable total effect ($X \rightarrow Y$), all paths in the flowchart, in which conditional independence is firmly established, unequivocally identify the causal chain from X through M to Y and hence full causal mediation as the underlying data generating mechanism. Paths that terminate inconclusively point to situations where claims about mediation rest on subjective prior assumptions about the underlying data generating mechanism (see also Table 2 in Section 4). However, even in inconclusive cases the suggested approach will result in less biased estimates of what may be a causal indirect effect, because it controls for measurement error and discretization of (potential) mediators.

The suggested methodology and procedural approach consists of interconnected steps and omitting one step may render the others ineffective. For instance, not accounting for measurement error may lead to a spuriously rejecting conditional independence, and thus a causally inconclusive result. What makes the proposed approach useful and novel is that it unifies the modeling of measurement error in and the discretization of observed indicators and the assessment of conditional independence between (latent) Y and X given (latent) M in a Bayesian framework.

Next we explain the procedure step-wise. We provide the details of Bayesian inference and Bayes factor computations in online Appendix G. Computer code with examples is available at [Github](#).

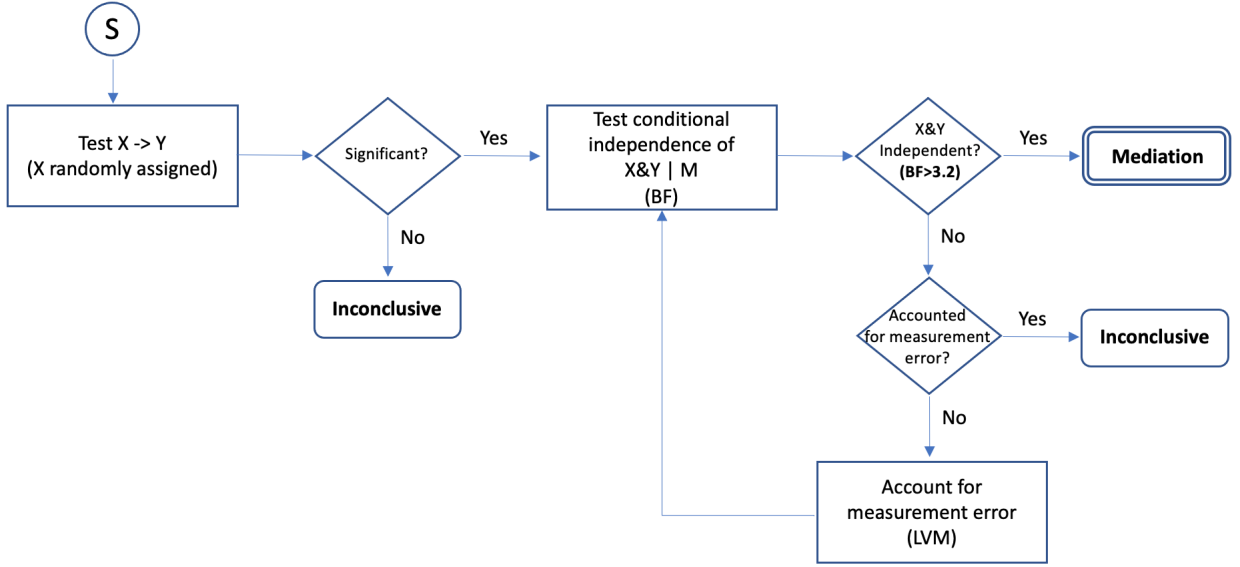


Figure 12: A flowchart of the proposed method for mediation analysis

- i. Estimate the simple mediation model.
- ii. Compute the Bayes factor (Equation 9) in the simple model.
- iii. If the BF strongly supports full mediation ($BF > 3.2$), proceed to step (vi), otherwise proceed to step (iv).
- iv. Estimate the latent variable model (LVM) with multiple indicators.
- v. Compute the Bayes factor in the LVM.
- vi. Report the Bayesian 95% PCI of the indirect and direct effects as well as the Bayes factor and the subjective prior settings.

We recommend using the reference prior for the direct effect and interpreting the Bayes factors based on Kass and Raftery (1995)’s guidelines (see Table 4). Our simulation study illustrates the robustness of the suggested approach across a wide range of possible settings pertaining to data generating mechanisms and estimation specifications. Moreover, in Online Appendix D, we provide guidelines for researchers who may wish to conduct further sensitivity analyses.

A Bayes factor in the interval $[0.32, 3.2]$ indicates that the information in the data is inconclusive about conditional independence. The obvious strategy then is to collect more data. Of course, it is possible to claim partial mediation with a $BF < 3.2$ and what

may be a significant indirect effect.²⁰ However, one should refrain from the claim to have tested mediation as the underlying causal model in these instances. In fact, the interpretation of the indirect effect as such entirely depends on prior knowledge in these cases. We also advise against making overconfident claims in favor/against conditional independence based on Bayes factors marginally larger than 3.2 or smaller than 0.32, and generally recommend sample sizes of $N = 350$ or larger in the search for the underlying causal structure in measurement-of-mediation designs. Moreover we offer simulation-based Bayesian power analysis in analogy to our simulation studies as part of the R package accompanying this paper.

Finally, strong evidence for conditional independence between Y and X given M in the simple model rules out *all* other possibilities, including measurement error in the mediator, but the causal chain from X through M to Y . In this case as instructed in step iii) of the procedure, we can right away confidently claim full mediation as the causal model generating the data (assuming evidence supporting total effect $\alpha > 0$, of course). However, the combination of latent variable modeling, MCMC estimation and Bayesian model comparisons facilitated by the latter enables researchers to harness the unique mapping from evidence for conditional independence between X and (latent) Y given (latent) M into evidence for mediation as the data generating causal model, even in the presence of measurement error.

We conclude by discussing potential extensions to the procedure suggested in this paper. First, our re-analysis of the study by [Tonietto and Malkoc \(2016\)](#) above shows how to leverage our suggested approach in the context of potential serial mediation. Similarly, we can use the suggested approach to obtain potentially conclusive results in the context of parallel mediation, in a causal sense. Evidence for parallel mediation as the underlying causal model results from conditional independence between X and Y given a set of mediators $\mathbf{M} = \{M_1, M_2, \dots\}$ where omitting any element from the set leaves X and Y conditionally dependent.

Second, when X not only changes the level of M but also how M transmits its causal effect to Y , sometimes referred to as a case of "moderated mediation" (see e.g., [Judd and Kenny, 1981](#); [Preacher et al., 2007](#)), even controlling M by experimental means does not shield Y from X . In this model, a specific, fixed M translates into a different Y depending

²⁰Recall that the reciprocal of the Bayes factor is a measure of evidence *against* conditional independence between Y and X given M in a model.

on X . Thus, establishing conditional independence between X and Y given M also rules out X as a moderator of the effect from M to Y , and more generally X as a conditioning argument to any causal consequences of M .²¹ While it is possible to extend our measure of data based evidence for $\beta_3 = 0$ to this case of moderated mediation, the lack of a direct effect in this model does not follow from a conditional independence relationship in the causal model. Finally, the generalization of our procedure to the more common case of manipulated third variable(s) Z , moderating the effects between X and M (β_1) or that between M and Y (β_2), or both involves testing for conditional independence under all levels of the moderating variable(s).

This paper is accompanied by computer code in package form available at [Github](#) that runs under R ([R Core Team, 2020](#)). The package includes functions that estimate parameters and Bayes factors measuring empirical support for conditional independence between X and Y given M , and accommodates multiple, potentially discretized measures of M and Y . In the latter case, Bayes factors measure evidence for conditional independence between X and latent Y given latent M . In this way, applications can benefit from the unique mapping from conditional independence to mediation as the causal data generating mechanism in so called measurement-of-mediation designs. Such designs, in which the putative mediator is only observed often suffer from measurement error or use discretized indicators. In addition, The package delivers exact small sample distributions of indirect effects. The examples included in the package illustrate how to assess sensitivity with respect to subjective prior assumptions, conduct a Bayes factor power analysis prior to data collection, and assess the convergence and Monte-Carlo stability of decision relevant quantities such as the Bayes factor.

²¹Establishing generalized conditional *mean* independence only of course does not rule out direct effects that are *exactly* mean preserving, which we argue to be practically irrelevant.

References

- Alter, A. L. and D. M. Oppenheimer (2008). Effects of fluency on psychological distance and mental construal (or why new york is a large city, but new york is a civilized jungle). *Psychological Science* 19(2), 161–167.
- Amrhein, V., S. Greenland, and B. McShane (2019). Scientists rise up against statistical significance.
- Andrews, M., X. Luo, Z. Fang, and A. Ghose (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science* 35(2), 218–233.
- Annis, J., J. G. Lenes, H. A. Westfall, A. H. Criss, and K. J. Malmberg (2015). The list-length effect does not discriminate between models of recognition memory. *Journal of Memory and Language* 85, 27–41.
- Baron, R. M. and D. A. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51(6), 1173.
- Berger, J., M. Draganska, and I. Simonson (2007). The influence of product variety on brand perception and choice. *Marketing Science* 26(4), 460–472.
- Bhargave, R., A. Mantonakis, and K. White (2016). The cue-of-the-cloud effect: When reminders of online information availability increase purchase intentions and choice. *Journal of Marketing Research* 53(5), 699–711.
- Bolger, N. (1998). Data analysis in social psychology. *Handbook of social psychology* 1, 233–65.
- Bollen, K. A. (1989). The consequences of measurement error. *Structural Equations with Latent Variables*, 151–178.
- Cheng, A. and C. Cryder (2018). Double mental discounting: When a single price promotion feels twice as nice. *Journal of Marketing Research* 55(2), 226–238.
- Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association* 90(432), 1313–1321.

- Cole, D. A. and S. E. Maxwell (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *Journal of abnormal psychology* 112(4), 558.
- Cole, D. A. and K. J. Preacher (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods* 19(2), 300.
- Dai, H., K. L. Milkman, and J. Riis (2015). Put your imperfections behind you: Temporal landmarks spur goal initiation when they signal new beginnings. *Psychological science* 26(12), 1927–1936.
- Dyachenko, T. L. and G. M. Allenby (2020). Is your sample truly mediating? a new way of thinking about mediation analysis.
- Ebbes, P. (2004). Latent instrumental variables. *A New Approach to Solve for Endogeneity, Ridderkerk: Labyrinth Publications*.
- Eddelbuettel, D., R. François, J. Allaire, K. Ushey, Q. Kou, N. Russel, J. Chambers, and D. Bates (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software* 40(8), 1–18.
- Feinberg, F. M. (2012). Mediation analysis and categorical variables: Some further frontiers. *Journal of Consumer Psychology* 22(4), 595–598.
- Ferraro, R., J. R. Bettman, and T. L. Chartrand (2008). The power of strangers: The effect of incidental consumer brand encounters on brand choice. *Journal of Consumer Research* 35(5), 729–741.
- Finch, J. F., S. G. West, and D. P. MacKinnon (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal* 4(2), 87–107.
- Franke, N., M. Schreier, and U. Kaiser (2010). The “i designed it myself” effect in mass customization. *Management science* 56(1), 125–140.

- Fritz, M. S., D. A. Kenny, and D. P. MacKinnon (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate behavioral research* 51(5), 681–697.
- Fritz, M. S. and D. P. MacKinnon (2007). Required sample size to detect the mediated effect. *Psychological science* 18(3), 233–239.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. CRC press.
- Glymour, C. N. (2001). *The mind’s arrows: Bayes nets and graphical causal models in psychology*. MIT press.
- Graf, L. K., S. Mayer, and J. R. Landwehr (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology* 28(3), 393–411.
- Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication monographs* 76(4), 408–420.
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of consulting and clinical psychology* 65(4), 599.
- Hoyle, R. H. (1999). *Statistical strategies for small sample research*. Sage.
- Iacobucci, D. (2008). *Mediation analysis*. Number 156. Sage.
- Iacobucci, D., N. Saldanha, and X. Deng (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of consumer psychology* 17(2), 139–153.
- Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological methods* 15(4), 309.
- James, L. R. and J. M. Brett (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology* 69(2), 307.
- James, L. R., S. A. Mulaik, and J. M. Brett (1982). *Causal analysis*. Beverly Hills: Sage.

- Judd, C. M. and D. A. Kenny (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation review* 5(5), 602–619.
- Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin* 64(5), 326.
- Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the american statistical association* 90(430), 773–795.
- Kenny, D. A. and C. M. Judd (2014). Power anomalies in testing mediation. *Psychological Science* 25(2), 334–339.
- Kleiman, T., C. Stern, and Y. Trope (2016). When the spatial and ideological collide: Metaphorical conflict shapes social perception. *Psychological science* 27(3), 375–383.
- Kramer, T., S. Spolter-Weisfeld, and M. Thakkar (2007). The effect of cultural orientation on consumer responses to personalization. *Marketing Science* 26(2), 246–258.
- Landwehr, J. R., B. Golla, and R. Reber (2017). Processing fluency: An inevitable side effect of evaluative conditioning. *Journal of Experimental Social Psychology* 70, 124–128.
- Landwehr, J. R., A. A. Labroo, and A. Herrmann (2011). Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science* 30(3), 416–429.
- Landwehr, J. R., D. Wentzel, and A. Herrmann (2013). Product design for the long run: Consumer responses to typical and atypical designs at different stages of exposure. *Journal of Marketing* 77(5), 92–107.
- Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*, Volume 711. John Wiley & Sons.
- Liu, X. and L. Wang (2021). The impact of measurement error and omitting confounders on statistical inference of mediation effects and tools for sensitivity analysis. *Psychological Methods* 26(3), 327.
- MacKinnon, D. (2012). *Introduction to statistical mediation analysis*. Routledge.

- MacKinnon, D. P., A. J. Fairchild, and M. S. Fritz (2007). Mediation analysis. *Annu. Rev. Psychol.* 58, 593–614.
- MacKinnon, D. P., J. L. Krull, and C. M. Lockwood (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention science* 1(4), 173–181.
- MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods* 7(1), 83.
- MacKinnon, D. P., C. M. Lockwood, and J. Williams (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research* 39(1), 99–128.
- Matzke, D., S. Nieuwenhuis, H. van Rijn, H. A. Slagter, M. W. van der Molen, and E.-J. Wagenmakers (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General* 144(1), e1.
- Muthén, B. and T. Asparouhov (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods* 17(3), 313.
- Muthén, B. and T. Asparouhov (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling: A Multidisciplinary Journal* 22(1), 12–23.
- Nook, E. C., S. F. Sasse, H. K. Lambert, K. A. McLaughlin, and L. H. Somerville (2018). The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological Science* 29(8), 1346–1357.
- Nuijten, M. B., R. Wetzels, D. Matzke, C. V. Dolan, and E.-J. Wagenmakers (2015). A default bayesian hypothesis test for mediation. *Behavior research methods* 47(1), 85–97.
- O’Hagan, A. and J. Forster (1994). Bayesian inference, volume 2b. *Kendall’s Advanced Theory of Statistics*.
- Packard, G. and J. Berger (2017). How language shapes word of mouth’s impact. *Journal of Marketing Research* 54(4), 572–588.

- Pearl, J. (2009). *Causality*. Cambridge university press.
- Pieters, R. (2017). Meaningful mediation analysis: Plausible causal inference and informative communication. *Journal of Consumer Research* 44(3), 692–716.
- Preacher, K. J. and A. F. Hayes (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior research methods, instruments, & computers* 36(4), 717–731.
- Preacher, K. J. and K. Kelley (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological methods* 16(2), 93.
- Preacher, K. J., D. D. Rucker, and A. F. Hayes (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate behavioral research* 42(1), 185–227.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rossi, P. E., G. M. Allenby, and R. McCulloch (2012). *Bayesian statistics and marketing*. John Wiley & Sons.
- Rouder, J. N. and J. Lu (2005). An introduction to bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic bulletin & review* 12(4), 573–604.
- Rouder, J. N., R. D. Morey, P. L. Speckman, and J. M. Province (2012). Default bayes factors for anova designs. *Journal of Mathematical Psychology* 56(5), 356–374.
- Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review* 16(2), 225–237.
- Rucker, D. D., K. J. Preacher, Z. L. Tormala, and R. E. Petty (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass* 5(6), 359–371.

- Schroeder, J., M. Kardas, and N. Epley (2017). The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological science* 28(12), 1745–1762.
- Sela, A. and R. A. LeBoeuf (2017). Comparison neglect in upgrade decisions. *Journal of Marketing Research* 54(4), 556–571.
- Shaddy, F. and A. Fishbach (2017). Seller beware: how bundling affects valuation. *Journal of Marketing Research* 54(5), 737–751.
- Shrout, P. E. and N. Bolger (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods* 7(4), 422.
- Song, H. and N. Schwarz (2009). If it’s difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science* 20(2), 135–138.
- Steenkamp, J.-B. E. and E. Fang (2011). The impact of economic contractions on the effectiveness of r&d and advertising: evidence from us companies spanning three decades. *Marketing Science* 30(4), 628–645.
- Sussman, A. B. and R. L. O’Brien (2016). Knowing when to spend: Unintended financial consequences of earmarking to encourage savings. *Journal of Marketing Research* 53(5), 790–803.
- Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82(398), 528–540.
- Thompson, D. V. and E. C. Ince (2013). When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research* 50(2), 228–240.
- Tingley, D., T. Yamamoto, K. Hirose, L. Keele, and K. Imai (2014). Mediation: R package for causal mediation analysis.
- Tofghi, D. and K. Kelley (2020). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence interval formation. *Multivariate behavioral research* 55(2), 188–210.

- Tonietto, G. N. and S. A. Malkoc (2016). The calendar mindset: scheduling takes the fun out and puts the work in. *Journal of Marketing Research* 53(6), 922–936.
- van Ravenzwaaij, D., G. Dutilh, and E.-J. Wagenmakers (2011). Cognitive model decomposition of the bart: Assessment and application. *Journal of Mathematical Psychology* 55(1), 94–105.
- Van Zant, A. B. and D. A. Moore (2015). Leaders’ use of moral justifications increases policy support. *Psychological science* 26(6), 934–943.
- Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the bayes factor. *Journal of Mathematical Psychology* 54(6), 491–498.
- Wagenmakers, E.-J., R. Wetzels, D. Borsboom, and H. L. Van Der Maas (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on bem (2011). *Journal of Personality and Social Psychology* 100(3), 426–432.
- Wedel, M. and C. Dong (2020). Banova: Bayesian analysis of experiments in consumer psychology. *Journal of Consumer Psychology* 30(1), 3–23.
- Yan, D. and A. V. Muthukrishnan (2014). Killing hope with good intentions: The effects of consolation prizes on preference for lottery promotions. *Journal of Marketing Research* 51(2), 198–204.
- Yuan, Y. and D. P. MacKinnon (2009). Bayesian mediation analysis. *Psychological methods* 14(4), 301.
- Zaval, L., E. M. Markowitz, and E. U. Weber (2015). How will i be remembered? conserving the environment for the sake of one’s legacy. *Psychological science* 26(2), 231–236.
- Zhang, J., M. Wedel, and R. Pieters (2009). Sales effects of attention to feature advertisements: a bayesian mediation analysis. *Journal of marketing research* 46(5), 669–681.
- Zhao, X., J. G. Lynch Jr, and Q. Chen (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of consumer research* 37(2), 197–206.

Online appendix

A Illustrative performance of existing approaches to causal mediation

A.1 Sensitivity analysis to the unobserved confounder

Imai et al. (2010) propose a sensitivity analysis of what may be a causal indirect effect to the strength of a possible error correlation ρ between the mediator and the outcome variable. If the indirect effect is robust to a range of reasonable ρ values, the analyst can be confident that mediator endogeneity is not an issue. Even though this method prevents misinterpreting correlation between M and Y as a causal M $\rightarrow$ Y effect, it assumes the causal direction between the putative mediator to the putative outcome variable.

To demonstrate, we simulate data (N=500) from the simple mediation model in Figure 2 with the following parameter values (see Equations 2 & 3 of the paper): $X \sim \text{Binomial}(0.5)$, $\beta_1 = 0.3$, $\beta_2 = 0.5$, $\beta_3 = 0$, $\text{Var}(U_M) = 0.2$, $\text{Var}(U_Y) = 0.2$, and conduct the indirect effect sensitivity analysis using the `mediation` package (Tingley et al., 2014) once with the correct specification and once misspecifying M as the outcome, and Y as the mediator.

The tabulated output "sens.out" in top panel of Figure A.1 shows the sensitivity analysis results for the correctly specified model. The output reports values of M-Y error correlation, for which the confidence interval of the indirect effect (ACME) includes 0. It shows that an error correlation of $\rho = 0.5$ in combination with an indirect effect equal to zero is an alternative rationale for this data. The "sens.out.misspec" output in the bottom panel of Figure A.1, displays ρ values for which the causally misspecified model yields insignificant indirect effects ($\rho = 0.4$ and $\rho = 0.5$). If the analyst deems $\rho \geq 0.4$ unlikely, she can interpret the estimated mediated effect causally in both cases, according to these results. That is, without prior knowledge about the causal ordering or measures variables, the analyst cannot distinguish the correct model from the misspecified model by comparing sensitivity analyses under different causal orderings.

```
> summary(sens.out)
```

Mediation Sensitivity Analysis for Average Causal Mediation Effect

Sensitivity Region

	Rho	ACME	95% CI Lower	95% CI Upper	R^2_M*R^2_Y*	R^2_M~R^2_Y~
[1,]	0.5	-0.0129	-0.0379	0.0121	0.25	0.1793

Rho at which ACME = 0: 0.5
R^2_M*R^2_Y* at which ACME = 0: 0.25
R^2_M~R^2_Y~ at which ACME = 0: 0.1793

```
> summary(sens.out.misspec)
```

Mediation Sensitivity Analysis for Average Causal Mediation Effect

Sensitivity Region

	Rho	ACME	95% CI Lower	95% CI Upper	R^2_M*R^2_Y*	R^2_M~R^2_Y~
[1,]	0.4	0.0105	-0.0011	0.0221	0.16	0.1148
[2,]	0.5	-0.0051	-0.0151	0.0048	0.25	0.1793

Rho at which ACME = 0: 0.5
R^2_M*R^2_Y* at which ACME = 0: 0.25
R^2_M~R^2_Y~ at which ACME = 0: 0.1793

Figure A.1: Results of the sensitivity analysis proposed by Imai et al. (2010) for the correctly specified model (top) and the misspecified model (bottom).

In fact plotting the indirect effect's confidence interval as a function of the mediator-outcome error correlation (ρ) in the misspecified model suggests that unless ρ is in a small range of values in proximity to 0.5, there is a causal effect from X to what is actually M, mediated by what is actually Y (Figure A.2).

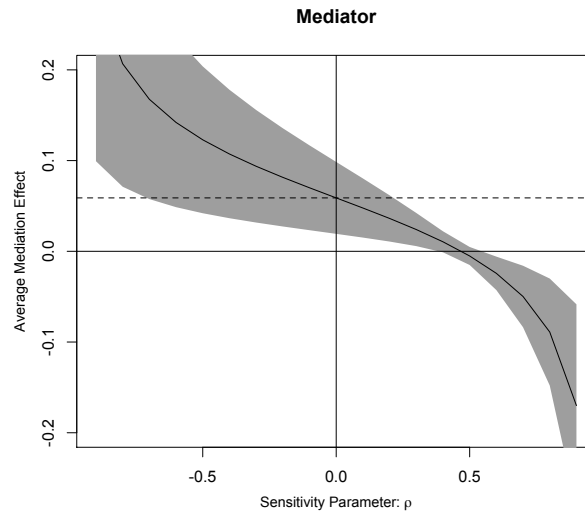


Figure A.2: Visualization of the indirect effect's sensitivity analysis to the "mediator"-outcome error correlation ρ .

Our proposed approach on the other hand clearly distinguishes between the correctly specified model (BF = 19.9), and the model where M is confused with Y (BF = 0.00012).

A.2 Latent Instrumental Variables

A potential way to measure causal relations in any model, including models of mediation is through instrumental variables. If there exist valid instruments for both the mediator and the outcome variable, exogenous variation in the instruments help identify the underlying causal model. However valid instruments are hard to find. Even if an instrument is found that is strongly correlated with the endogenous variable, there is no way to empirically establish the required exclusion restriction. [Zhang et al. \(2009\)](#) propose to rely on the latent instrumental variables (LIV) approach introduced by [Ebbes \(2004\)](#) to account for the possible endogeneity of the observational variables (in our case only the mediator) in mediation analysis. The latent instrument distinguishes between exogenous variation in the mediator and hence in the outcome due to the mediator, and variation due to unobserved confounders or measurement error.

However, just as [Imai et al. \(2010\)](#)’s sensitivity analysis, the application of LIV in this context requires conditioning on the correct order of measured variables. We again illustrate numerically. We simulate data from the simple mediation model with parameter values: $X \sim \text{Binomial}(0.5)$, $\beta_1 = 1$, $\beta_2 = 3$, $\beta_3 = 0$, $\text{Var}(U_M) = 1$, $\text{Var}(U_Y) = 1$. We then use the `REndo` package in R to estimate β_2 using LIV, once in the correct model and once in a model where M and Y are switched.²² Figure A.3 shows the summary of the output of the regression of Y on the instrumented M. The LIV method estimates β_2 accurately in the correct model. Note that M here is in fact generated by an instrument (X).

However, the LIV also infers a significant $Y \rightarrow M$ effect, even though such an effect does not exist. The output of the regression of M on an LIV instrumented Y in Figure A.4 shows a significant effect from Y on M ($\beta_Y = 0.33$).

In fact, unless Y itself is generated by an additional instrument to be captured by the LIV, LIV cannot identify the causal direction. Here again, the Bayes factor of the correctly ordered model shows substantial evidence in favor of full causal mediation (BF = 7.2), and the Bayes factor of the misspecified model shows decisive evidence against conditional independence (BF = $1.8 * 10^{-13}$).

²²Since $\beta_3=0$, regressing Y on M yields an unbiased estimate of β_2 .

```

Call:
latentIV(formula = Y ~ M, data = as.data.frame(Data), start.params = c('(Intercept)' = 1,
  M = 3))

Coefficients:
              Estimate Std. Error z-score Pr(>|z|)
(Intercept)  0.92615    0.28914   3.203   0.0014 **
M             3.02063    0.09617  31.410   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Further parameters estimated during model fitting:
      pi1      pi2  theta5  theta6  theta7  theta8
2.09123  3.96488  0.53150  1.02547 -0.02804  1.05134
(see help file for details)

Initial parameter values:
(Intercept)=1 M=3 pi1=2.9959 pi2=4.402 theta5=0.5 theta6=1
theta7=0.5 theta8=1

The value of the log-likelihood function: 3192.358
AIC: -6368.716 , BIC: -6329.454
KKT1: FALSE KKT2: TRUE Optimx Convergence Code: 0

```

Figure A.3: The LIV estimate of the effect of M on Y in the correctly specified model.

```

Call:
latentIV(formula = M ~ Y, data = as.data.frame(Data), start.params = c('(Intercept)' = 1,
  Y = 1))

Coefficients:
              Estimate Std. Error z-score Pr(>|z|)
(Intercept) -0.329998    0.073518  -4.489   8e-06 ***
Y             0.334068    0.007361  45.382   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Further parameters estimated during model fitting:
      pi1      pi2  theta5  theta6  theta7  theta8
6.4646 12.6825  0.4471  0.1162 -0.3664  8.7235
(see help file for details)

Initial parameter values:
(Intercept)=1 Y=1 pi1=9.9575 pi2=14.3223 theta5=0.5 theta6=1
theta7=0.5 theta8=1

The value of the log-likelihood function: 3193.176
AIC: -6370.352 , BIC: -6331.09
KKT1: FALSE KKT2: TRUE Optimx Convergence Code: 0

```

Figure A.4: The LIV estimate of the effect of M on Y in the misspecified model.

A.3 Heterogeneous Mediation Model

Based on a structural heterogeneity argument, [Dyachenko and Allenby \(2020\)](#) model the heterogeneity of the mediated effect through a Bayesian mixture model. The model categorizes respondents into a group where the effect is fully mediated, and a group where it is not. The model estimates the probability of being assigned to the full mediation group, which could, but as we will show should not, be interpreted as the probability of full mediation (i.e., conditional independence). The reason is that the model groups

individuals into the full mediation group based on fit *relative to other respondents*. This will be problematic in the case of two groups, one of which is "further away" from full mediation than the other but neither conforms to full mediation. Next, we illustrate this problem numerically.

In the first example we generate a sample ($N=1000$) with two partially mediated groups, where the second group (Group2) exhibits a stronger *direct* effect than the other: Group1 (90% of the observations): $X \sim N(0, 1), \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.3, Var(U_M) = 1, Var(U_Y) = 1$.

Group2 (10% of the observations): $X \sim N(0, 1), \beta_1 = 1, \beta_2 = 1, \beta_3 = 3, Var(U_M) = 1, Var(U_Y) = 1$.

We estimate a simplified version of the mixture model proposed by [Dyachenko and Allenby \(2020\)](#) where an indicator ind_i determines to which component (full/partial mediation) individual i belongs. The indicator ind_i is distributed Bernoulli with hierarchical prior parameter $pmed$. The latter could be interpreted as the probability of a fully mediated treatment, and could potentially be read as the amount of evidence supporting full mediation in the data.

In reality however, the mixture model categorizes the observations from the group with weaker direct effect as fully mediated simply because the full mediation model fits these observations better than *observations from the group* with stronger direct effect, and the statistical distance between observations from different groups is larger than that between observations from the first group and the full mediation model. Table A.1 shows the estimation results, where the probability of full mediation in the group is estimated to be $\widehat{pmed} = 0.86$, whereas the actual probability of full mediation in the population is zero. Applying our procedure yields decisive evidence against conditional independence in both groups ($BF_{Group1} = 1.76 \times 10^{-7}$ and $BF_{Group2} = 4.63 \times 10^{-94}$).

Parameters→ Mediation Group↓	$\widehat{pmed}$	$\hat{\beta}_2$	$\hat{\beta}_3$
<i>Full</i>	0.86	0.67	0
<i>Partial</i>	0.14	0.77	3.35

Table A.1: Estimates of the heterogeneous simple mediation model for a data with two partially mediated groups.

In the second example we change the data generating mechanism in the first group (Group1) such that there is no mediation in this group. Both the direct and the indirect

effects are zero in this group. However, an unobserved confounder introduces dependence between M and Y. We generate a total of N=1000 observations from the two models, in the first, M and Y have no causal link but have correlated error terms due to an unobserved variable. The other group of observations are generated from a partial mediation model as in the first example's Group2. The data generating values are as follows:

Group1 (90% of the observations): $X \sim N(0, 1), \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, Var(U_M) = 1, Var(U_Y) = 1, Cov(U_M, U_Y) = -.5$.

Group2 (10% of the observations): $X \sim N(0, 1), \beta_1 = 1, \beta_2 = 1, \beta_3 = 3, Var(U_M) = 1, Var(U_Y) = 1$.

Although the correlated error term in Group1 induces a spurious direct effect, this effect is small *relative to Group2's direct effect* ($\beta_3 = 3$). Consequently, observations from Group1 tend to be classified into the "full mediation" part of the mixture (Table A.2).

Parameters→ Mediation Group↓	$\widehat{pmed}$	$\hat{\beta}_2$	$\hat{\beta}_3$
<i>Full</i>	0.86	-0.2	0
<i>Partial</i>	0.14	0.3	2.86

Table A.2: Estimates of the heterogeneous simple mediation model for a group with no mediation and a partially mediated group.

Again, Bayes factors computed for both simulated groups show decisive evidence against full mediation ($BF_{Group1} = 7.7 \times 10^{-17}$ and $BF_{Group2} = 1.47 \times 10^{-14}$).

B Identification of the measurement model

In line with [Bollen \(1989\)](#)'s 2-indicator rule, we need at least 2 indicators for each latent variable to identify the measurement model. There are a range of parameter values for the single indicator measurement model shown in Figure B.1 that can rationalize a single data set equally well. This range includes models with zero and non-zero direct effects, which are consistent with full mediation and partial mediation respectively. To formally illustrate this identification problem, we analyze covariances and derive how a model with non-zero direct effect might imply the same covariance matrix as the full mediation model with measurement error.

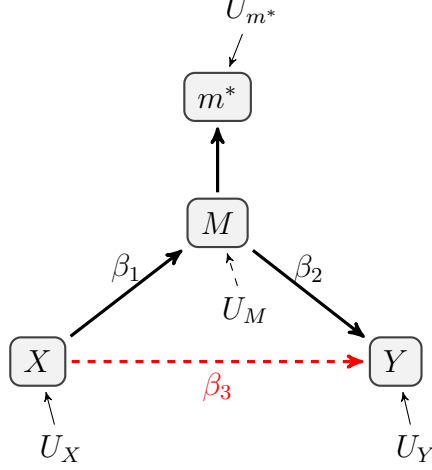


Figure B.1: Direct effect counterbalancing the latent mediator error (directed acyclic graph).

Intuitively, the effect from X to m^* (via unobserved M) fits $Cov(X, m^*)$, the product of effects from X to M and from M to Y rationalizes $Cov(X, Y)$. Finally, $Cov(m^*, Y)$ is a function of the former effects and increasing in $Var(U_M)$. Therefore, to the extent that $Var(U_{m^*})$ remains positive, $Var(U_M)$ can adjust to allow for a (hypothetical) direct effect from X to Y .

To illustrate this identification problem, consider a dataset with $N = 10000$, generated according to the model in Figure B.1 with the following parameter values : $X \sim U(0, 1)$, $\beta_1 = 1$, $\beta_2 = 3$, $\beta_3 = 0$, $Var(U_M) = 1$, $Var(U_Y) = 3$, and $Var(U_{m^*}) = 2$. Figure B.2 (left panel) plots measurement error variance values in the interval $[0, Var(U_{m^*}) + Var(U_M)]$ against corresponding values of the model's joint likelihood function $p(X, m^*, Y | \beta_1, \beta_2, \beta_3, Var(U_M), Var(U_Y), Var(U_{m^*}))$. The data generating parameter values imply full mediation with measurement error. However Figure B.2 (right panel), which plots the parameter values associated with likelihood values in the left panel, shows that for each measurement error variance in the range of $Var(U_{m^*}) \in (0, 3]$, there are corresponding values of other parameters that yield the same covariance matrix and hence equal likelihood values. The parameters adjust within the boundaries where error variances are non-negative so the set of β_3 values that yields maximum likelihood has its upper bound at $Var(U_{m^*}) = 0$, and its lower bound at $Var(U_Y) = 0$, and the non-negativity constraint on $Var(U_M)$ does not become binding here.

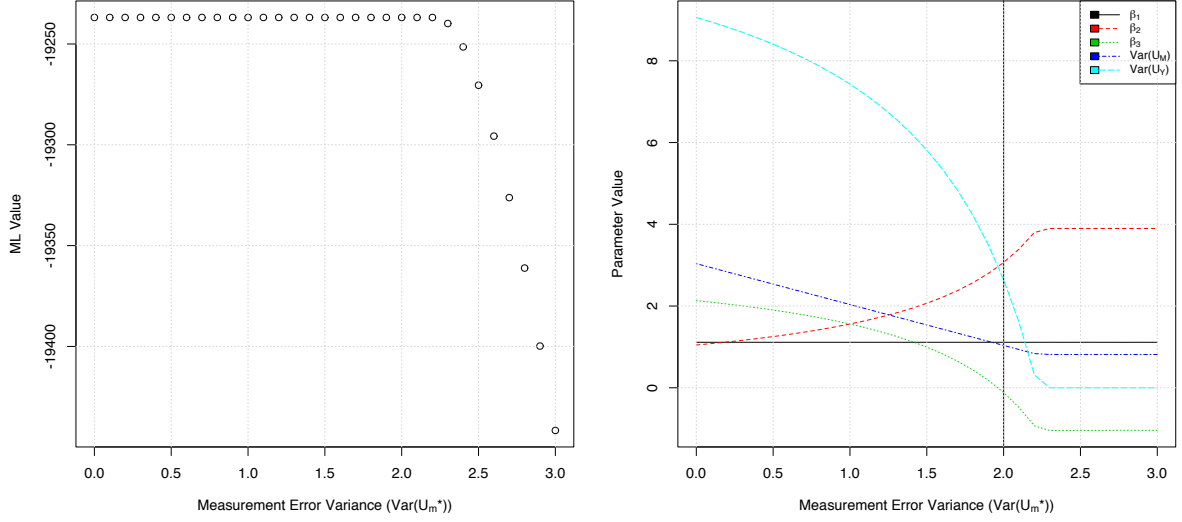


Figure B.2: Likelihood profile (left) and its corresponding parameter values of the unidentified measurement error model with direct effect. Vertical line on the right pane indicates true values.

In short, as long as the variance of the latent mediator error term ($Var(U_M)$) is free to adjust itself to changes in the direct effect, the latent variable model cannot be point-identified. A model with more than one indicator for the latent mediator does not have this problem. Intuitively, in such a model $Var(U_M)$ cannot freely adjust to balance out the direct effect anymore, because it is essentially a function of variances *and* covariances of indicator variables ($m_1^*, m_2^*, \dots$), which are observed fixed values.

We illustrate this circumstance comparing the joint posterior from estimation of an identified and an unidentified model. We generate data with 2 indicators for the mediator with the following parameter values: $X \sim U(0, 1)$, $\beta_1 = 1$, $\beta_2 = 3$, $\beta_3 = 0$, $Var(U_M) = 1$, $Var(U_Y) = 1$, $Var(U_{m_1^*}) = 1$, $Var(U_{m_2^*}) = 2$, $Var(U_{y_1^*}) = 2$, $Var(U_{y_2^*}) = 1$, $\{\lambda_{20}, \lambda_{21}\} = \{1, 1.5\}$, and $\{\tau_{20}, \tau_{21}\} = \{1, 2\}$. We then estimate a multiple indicator, i.e., a latent variable model using indicators for both latent M and Y, and a single indicator model using composite measures obtained by taking the average of the indicators. In this single indicator model the averages of the indicators are taken to equal M and Y.

Figure B.3 shows the joint posterior of $Var(U_M)$ and the direct effect β_3 for the single indicator model (left panel) and the multiple indicator model (right panel). In the single indicator model the two variables, as discussed earlier, are highly correlated. In fact, the only reason that the joint posterior distribution of β_3 and $Var(U_M)$ does not collapse to a perfect line are the proper subjective prior distributions for these parameters. The

correlation vanishes in the multiple indicator model. As multiple indicators prevent the perfect likelihood trade-off between β_3 and $Var(U_M)$, the model jointly point identifies the direct effect β_3 and the exogenous variance in the mediator M, $Var(U_M)$.

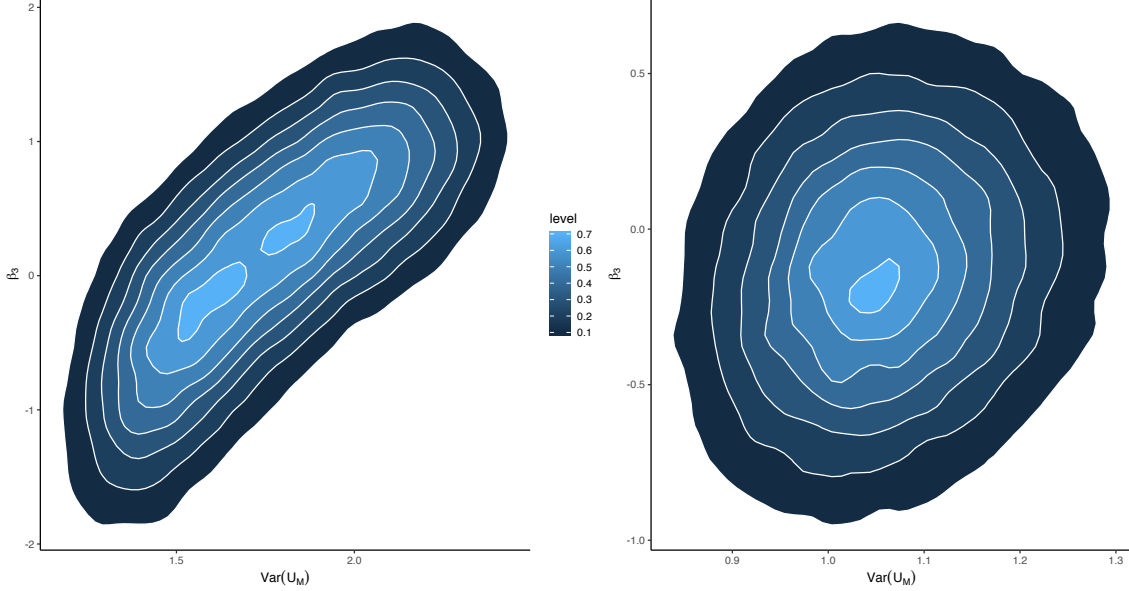


Figure B.3: Comparison of joint posterior distributions of latent mediator's error variance ($Var(U_M)$) and the direct effect(β_3) in the single indicator (left panel) and the multiple indicator (right panel) models

C Stable computation of Bayes factor

Bayes factor computation can be numerically unstable as both the denominator and the numerator of Equation 9, i.e., the marginal likelihoods of the models being compared, are integrals without a closed form that need to be computed using numerical methods. Because we are interested in a nested-model comparison, we can take advantage of the Savage-Dickey ratio, which is derived as follows (see also [Rossi et al., 2012](#)):

We start with the Bayes factor ratio as in Equation 9, where M_0 is the mediation model with parameters $\{\beta_3, \Theta\}$, where $\beta_3 = 0$, and Θ are all the parameters in the mediation model in addition to β_3 . For M_1 , β_3 is unrestricted. Then we have:

$$BF_{01} = \frac{p(y|M_0)}{p(y|M_1)} = \frac{\int \ell(\Theta|y, m, X)p(\Theta)d\Theta}{\int \int \ell(\Theta, \beta_3|y, m, X)p(\Theta, \beta_3)d\Theta d\beta_3} \quad (10)$$

Where $\ell(\bullet|y, m, X)$ denotes the likelihood of the data (y, m, X) that may include multiple indicators of Y and M , i.e., $\{y\}$ and $\{m\}$, to overcome measurement error as conditioning arguments to the posterior. Noting that $\ell(\Theta|y, m, X) = \ell(\Theta, \beta_3|y, m, X)|_{\beta_3=0}$,

and specifying the conditional prior $P(\Theta|\beta_3 = 0) = \frac{p(\Theta, \beta_3)}{\int p(\Theta, \beta_3) d\Theta} \Big|_{\beta_3=0}$ for M_0 , we can derive the following expression from Equation 10:

$$\frac{\int \ell(\Theta, \beta_3|y, m, X) p(\Theta, \beta_3) d\Theta}{p(\beta_3) \int \int \ell(\Theta, \beta_3|y, m, X) p(\Theta, \beta_3) d\Theta d\beta_3} \Big|_{\beta_3=0} = \frac{\int p(\Theta, \beta_3|y, m, X) d\Theta}{p(\beta_3)} \Big|_{\beta_3=0} \quad (11)$$

The Savage-Dickey ratio in Equation 11 shows that the Bayes factor comparing a nested constrained model to its unconstrained counterpart is equal to the ratio of the marginal posterior over the marginal prior of constrained parameters in the unconstrained model, both evaluated at the constraint. In our application, the target parameter is β_3 and the constraint is $\beta_3 = 0$.

Parameters in Θ need to be integrated out to obtain the marginal posterior of β_3 . At first, the need to compute this integral for the LVM we introduced in this paper may seem like a major obstacle. However, based on $r = 1, \dots, R$ draws from the joint posterior distribution $p(\beta_3, \Theta|y, m, X)$, we can approximate Equation 11 to an arbitrary degree of precision using Equation 12, provided that at least the full conditional posterior of β_3 is available in closed form.

$$BF_{01} \approx \frac{1/R \sum_{r=1}^R p(\beta_3|y, m, X, \Theta^r)}{p(\beta_3)} \Big|_{\beta_3=0} \quad (12)$$

The moments (mean and variance) indexing the full conditional posterior distribution in the numerator of Equation 12 are obtained at no additional computational cost as by-products of the MCMC algorithm following [Chib \(1995\)](#).

We illustrate the stability of our Bayes factor computation and the numerical consistency of the computed Bayes factors using one of the re-analyzed published studies from Section 5 ([Sussman and O'brien, 2016](#), study 3) as example. The target quantity here are simulated numerical standard errors of Bayes factors as a function of the MCMC length (R).

The numerical standard errors are computed as the standard deviation of Bayes factor distributions obtained from repeated, independent MCMC runs with different lengths. We use 100 runs for each distribution, all with a burn-in sample of 5000 draws. The distributions, plotted in Figure C.1, become more concentrated as R increases, as reflected by the change in scaling of the x-axis. Red vertical lines designate the mean of the

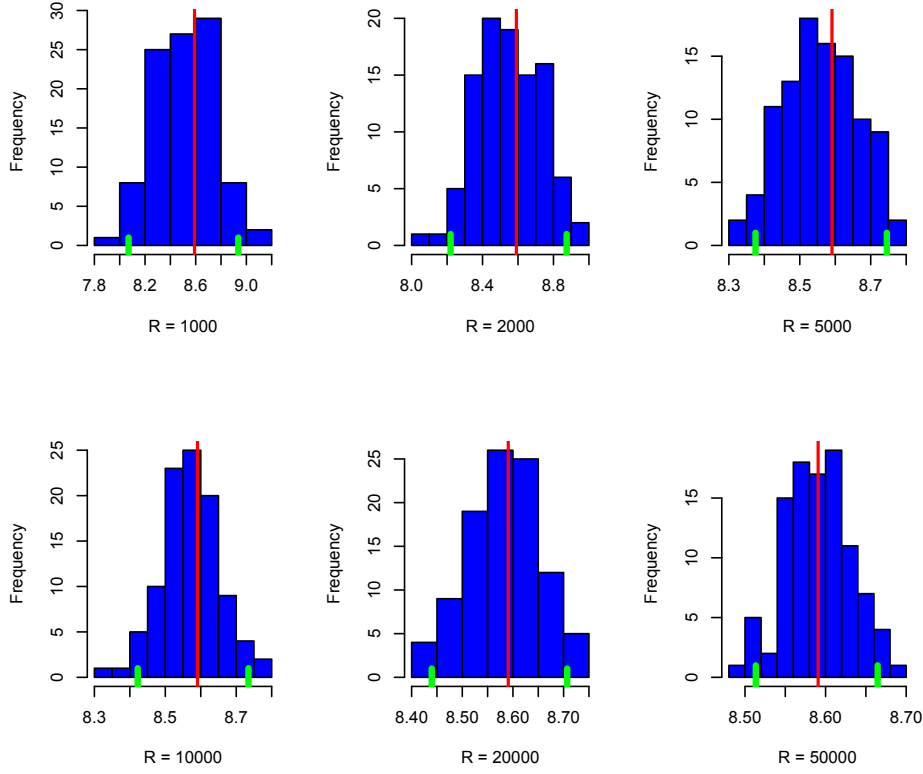


Figure C.1: Distributions of Bayes factors obtained from repeated MCMC runs with different lengths. [Sussman and O'brien \(2016\)](#), Study 3.

most concentrated distribution (corresponding to $R=50000$), which is well inside the 95% confidence interval (marked by green vertical lines) of all Bayes factor distributions obtained by independently repeating MCMC chains with fewer draws.

Standard deviations of these distributions, i.e., numerical standard errors presented in Table C.1 again show that the distributions become narrower and the accuracy of the estimates increases as we allow for longer MCMC runs. These results testify to the numerical consistency and reliability of our Bayes factor computations. Numerical standard errors of Bayes factors for all the re-analyzed studies, computed from MCMC runs with $R=10000$, are in the last column of Table F.1.

Implementation of the procedure in `Rcpp` facilitates short runtimes. For instance estimation of a latent variable model and Bayes factor computation for [Sussman and O'brien \(2016\)](#)'s study 3 ($N=408$) with an MCMC of length $R=50000$ take 27.96 seconds in total on a standard PC. Therefore, long MCMC runs for reliable, numerically accurate Bayes factors are highly practical for applied research.

MCMC iterations (R)	1000	2000	5000	10000	20000	50000
BF numerical stdandard error	0.23	0.18	0.1	0.08	0.07	0.04

Table C.1: Bayes factor consistency in the length of the MCMC chain for the mediation study in [Sussman and O'brien \(2016\)](#).

One could consider the Deviance Information Criterion (DIC) (e.g., [Gelman et al., 2013](#), pp 180-184) as an alternative measure to Bayes factors for model comparison. The DIC measure proposed by [Gelman et al. \(2013\)](#) can be computed for the simple mediation model as in Equation 13, where θ denotes all the parameters of the model, and the second term is the variance of the likelihood evaluated at posterior draws of θ .

$$DIC = -2 \sum_{r=1}^R \log(P(y, m, X|\theta_r)) + 2var_{\theta|y,m,X}[\log(P(y, m, X|\theta))] \quad (13)$$

The first term in Equation 13, the average log-likelihood of the data given the posterior, is a goodness of fit measure, whereas the second term, the variance of the log-likelihood over the posterior, is a penalty term for model complexity. The difficulty with this measure, when computing evidence in favor of a model with conditional independence, is that under conditional independence both models M_0 and M_1 converge to the same fit in the limit of increasing data information. And in this limit, the additional complexity from having β_3 in the model does not matter anymore.

Next, we illustrate the resulting inconsistency of the DIC numerically. We simulate data as follows: $X \sim N(0, 1)$, $\beta_1 = 0.3$, $\beta_2 = 0.5$, $\beta_3 = 0$, $Var(U_M) = 0.2$, $Var(U_Y) = 0.2$. Then, we compare the distribution of DIC differences with the distribution of Bayes factor values both obtained from 100 datasets (N=1000) in Figure C.2. Whereas the Bayes factor distribution has its largest mass over values that indicate strong evidence supporting full causal mediation ($BF_{95\%CI} = [6, 61]$), DIC values are insignificant ($DIC_{95\%CI} = [-2.3, 2.4]$) and the mode of the DIC distribution is roughly at -2 which is equal to the punishment of the measure for the extra parameter of the model without conditional independence (β_3).

Increasing the sample size does not affect the evidence supporting the true model according to the DIC. Figure C.3 compares Bayes factor and DIC distributions for larger samples (N=10000).

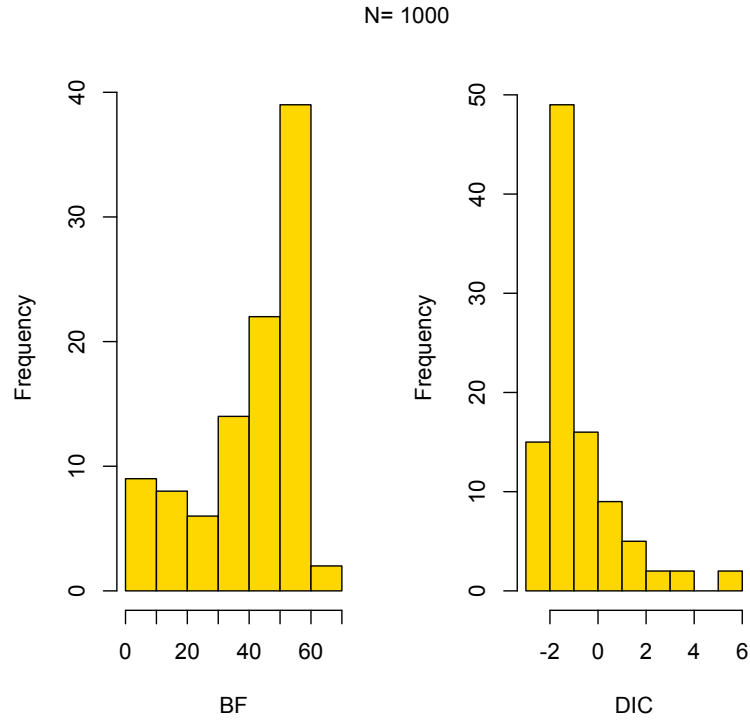


Figure C.2: Comparison of distributions of DIC and Bayes factor values obtained from 100 datasets with $N=1000$.

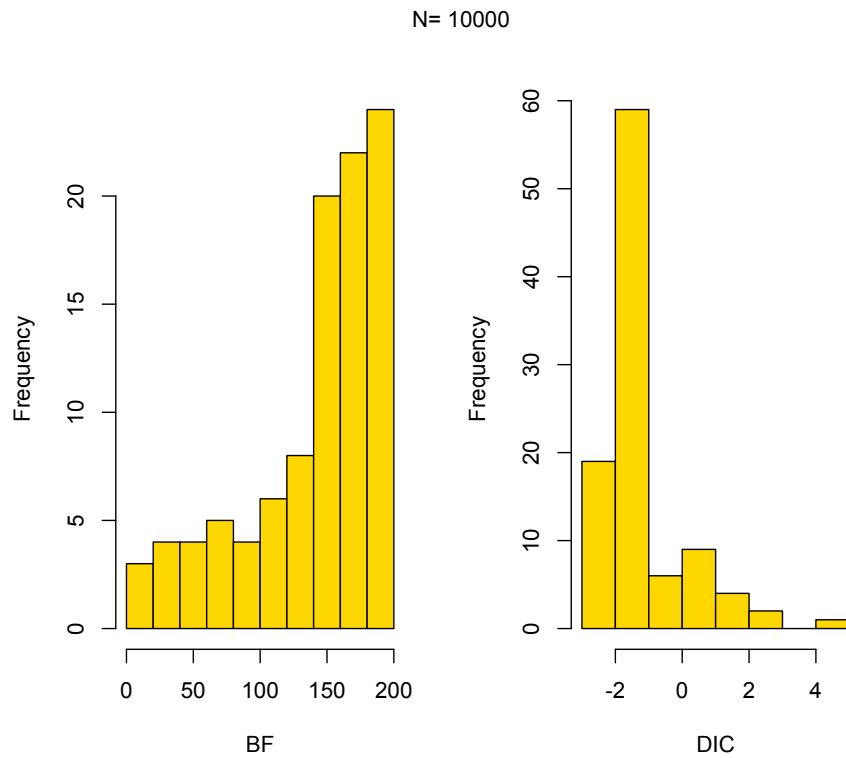


Figure C.3: Comparison of distributions of DIC and Bayes factor values obtained from 100 datasets with $N=10000$.

DIC values remain insignificant ($DIC_{95\%CI} = [-2.3, 2.3]$). In contrast, the Bayes factor distribution ($BF_{95\%CI} = [24, 185]$) has its mode at values indicating decisive evidence in favor of full causal mediation (>100), reflecting the increase in data-based evidence supporting conditional independence.

In addition to the consistency problem described above, computing DIC for latent variable models requires likelihood evaluations that would have to be simulated in a context where MCMC based inference leverages data augmentation.

D Illustrative sensitivity analyses

D.I Bayes factor sensitivity analysis

Here we illustrate an analysis of the sensitivity of the Bayes factor to the choice of subjective prior for β_3 . Specifically, we vary the prior variance in $\beta_3 \sim N(0, \sigma^2)$ on a grid, and plot it against the resulting Bayes factor measuring support for $\beta_3 = 0$. A useful way to interpret the results from the sensitivity analysis is as follows: if the evidence threshold of 3.2 (or a subjectively required larger value, see Table 4 in the paper) is only crossed once the prior is so diffuse (large σ^2) to likely support posterior estimates of β_3 that a researcher considers outrageously large a priori, the data fail to support $\beta_3 = 0$ and thus conditional independence.²³ If, in contrast, the evidence threshold is already crossed at priors that maybe considered informatively concentrated around zero (small σ^2) in the substantive domain under study, the data based evidence for $\beta_3 = 0$ and thus conditional independence is strong. More specifically, it is strong, in the eyes of all researchers that do not strongly believe a priori in even smaller direct effects, to who the prior with the small σ^2 is still overly diffuse.

Figure D.1 summarizes sensitivity with respect to the subjective prior for β_3 for two published studies. The left panel investigates the first mediation study in [Zaval et al. \(2015\)](#). It shows that even with a prior variance as small as $\sigma^2 = .5$, the Bayes factor points to substantial evidence supporting $\beta_3 = 0$ and hence (full) mediation. Therefore, we can be confident of our result in this case, unless we strongly believe in the presence of tiny direct effects a priori, as explained above.

²³At least to the group of researchers that agrees about the unduly diffuse nature of the subjective prior at this point.

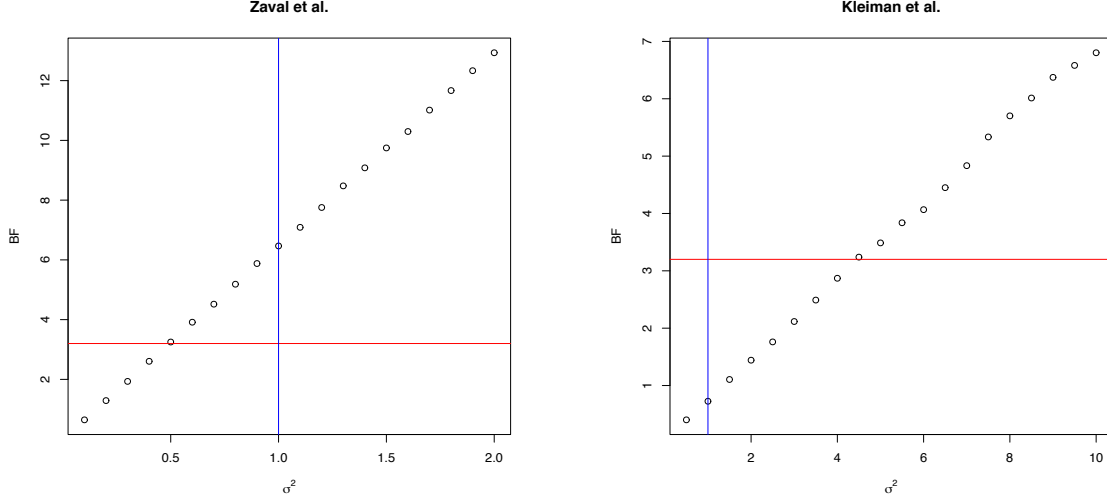


Figure D.1: Bayes factor sensitivity to direct effect's prior variance in [Zaval et al. \(2015\)](#) (left panel) and in [Kleiman et al. \(2016\)](#) (right panel). The red horizontal line indicates the lower bound for substantial evidence supporting (full) mediation. The blue vertical line indicates the default prior variance.

The right panel investigates the mediation study in [Kleiman et al. \(2016\)](#). This plot shows that for the Bayes factor to substantially support $\beta_3 = 0$, we need a very diffuse prior ($\sigma^2 = 4.5$). Based on this result, all researchers who did not expect (absolutely) large direct effects a priori and thus consider $\sigma^2 = 4.5$ to be informatively concentrated around zero, will confidently reject conditional independence. While such sensitivity analyses are extremely useful when different researchers cannot agree on the basic characterization of a given subjective prior, we keep with the reference prior $\beta_3 \sim N(0, 1)$ in our analyses, and supported by our survey of recently published estimates of β_3 (see Figure 7 in the paper).

D.II Assessing the direct effect prior in the multiple indicator mediation model with categorical observables

Latent variable models pose additional challenges to assessing a reasonable prior for β_3 . When multiple indicators are only observed on discretized, ordinal scales, we capture the potentially non-linear relationship between observed scale categories and continuous latent variables through freely (up to the order and a fixed origin) estimated indicator specific cut-points (see Equation 8). Scale identification of latent M and Y is achieved by fixing the corresponding (conditional) variance terms and parameters $\{\lambda_{i1}\}_{i=1}^K$ (for M), and $\{\tau_{i1}\}_{i=1}^L$ (for Y). As such, it is natural to assess the subjective prior for β_3 through the prior distribution of variance explained in latent Y by X via the direct effect β_3 . A

prior that is too diffuse, i.e., admits unreasonably large β_3 (in an absolute sense) with high probability will induce a distribution of variance explained concentrated at 1.

Variance explained is of course a function of the variance in X and the variance explained through the indirect effect. However, it is coherent to assess the prior for β_3 for the purpose of model comparison based on the achieved variance in X and posterior knowledge of β_1 and β_2 , maybe even gleaned from maximum likelihood estimation. The goal is again to calibrate a prior expressing beliefs about the strength of a direct effect, now in relative terms (variance explained) that is concentrated away from unrealistically large values to avoid the trivial mapping from an overly diffuse prior onto the simpler model.

Figure D.2 illustrates different *prior* distributions of the percentage of variance explained through the direct effect given different prior variances for the direct effect, fixed indirect effects: $\beta_1 = \beta_2 = 1.5$, and X distributed uniformly. These distributions show how more diffuse (higher variance) direct effect priors more likely translate into unreasonably large prior variance explained through the direct effect with high probability (see e.g., the bottom right panel in Figure D.2 with $\beta_3 \sim N(0, 100)$). As previously noted in Section 4 of the paper, it is coherent to assess the prior for β_3 even conditional on data based guesses for β_1 and β_2 . However, coherence requires a precommitment to *a priori* reasonable percentages of variance explained in latent Y through a possible direct effect. In practice, we recommend using the suggested reference prior (see the top right panel in Figure D.2), coupled with an assessment of the *prior* percentage of variance explained in latent Y through the direct effect conditional on posterior mean estimates of β_1 and β_2 . By the choice of variance normalization for identification, the reference prior will concentrate prior percentages of variance explained through the direct effect away from 1. However, reporting the *prior* distribution of percentage of variance explained through the direct effect may help inform or convince third parties of the necessary prior concentration for the Bayes factor to convey data based evidence.

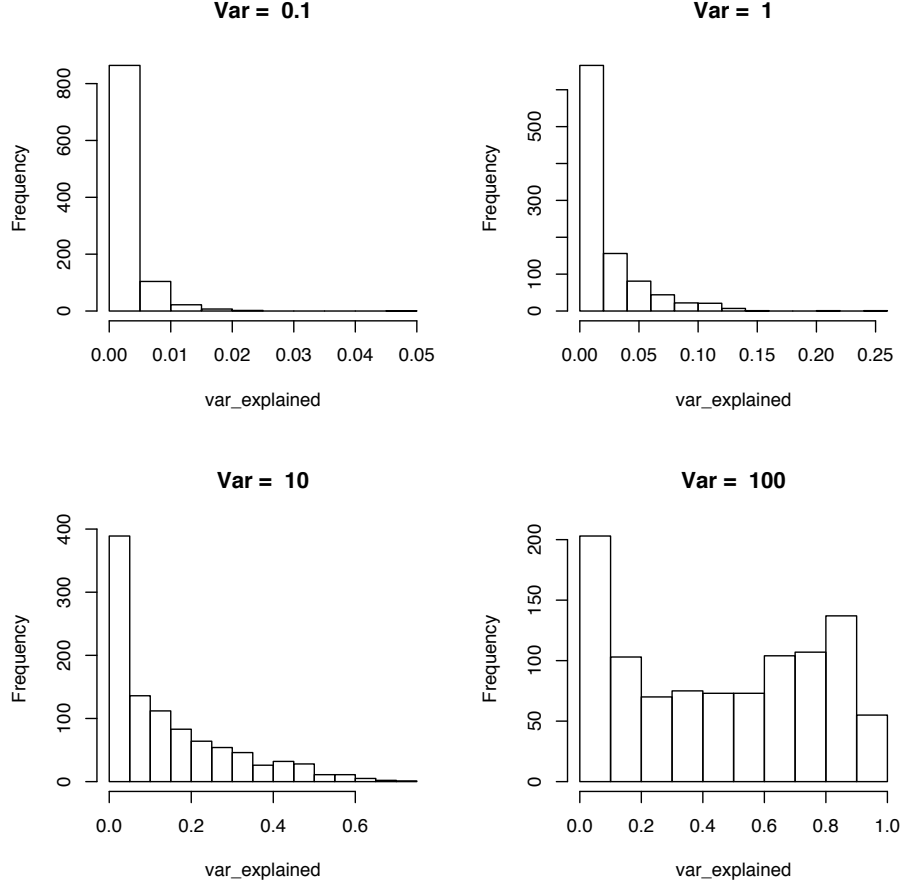


Figure D.2: Sensitivity of the percentage of a priori variance explained through the direct effect to the direct effect's prior variance in the categorical multiple indicator model.

E Simulation study

Table E.1-E.9 detail the results of the simulation study we conducted to examine the robustness of the procedure when applied under various conditions and specifications. Each table cell corresponds to 100 datasets simulated from a model specification with certain parameter values. Each cell contains the mode of the distribution of Bayes factors (tables E.1, E.4, and E.7), and percentage of datasets where Bayes factor correctly identifies or rejects conditional independence (tables E.2, E.5, and E.8). To compare Bayes factor results with the results from the approaches based on the indirect effect estimates, tables E.3, E.6, and E.9 show the percentage of datasets with significant indirect effect.

3 sets of specifications are studied: 1- datasets containing mediation, 2- datasets without mediation but an M-Y unobserved confounder, 3- datasets containing mediation but where mediator and outcome variables are confused with each other by the analyst

(M-Y directional misspecification).²⁴ For each of the three model specifications we generate datasets that vary across different dimensions:

(a) Indirect and direct effect values: We follow [Tofighi and Kelley \(2020\)](#) in choosing the indirect effect ($\beta_1 \cdot \beta_2$) and direct effect (β_3) values in the simulation. These values are in line with the range of the estimated effects in the re-analysis of the published studies.

(b) Measurement error: We examine four different conditions with respect to measurement error variances for mediator and outcomes indicators. 1. LL: $\sigma_{m_i^*} = \sigma_{y_j^*} = 0.05$, $i = 1, \dots, M_{ind}, j = 1, \dots, Y_{ind}$, 2. LH: $\sigma_{m_i^*} = 0.05$, $\sigma_{y_j^*} = 0.5$ $i = 1, \dots, M_{ind}, j = 1, \dots, Y_{ind}$, 3. HL: $\sigma_{m_i^*} = 0.5$, $\sigma_{y_j^*} = 0.05$ $i = 1, \dots, M_{ind}, j = 1, \dots, Y_{ind}$, 4. HH: $\sigma_{m_i^*} = \sigma_{y_j^*} = 0.5$, $i = 1, \dots, M_{ind}, j = 1, \dots, Y_{ind}$. We choose these measurement error variances based on their resulting reliability levels. The high measurement error variance (H) results in an item reliability = 0.6, and the low measurement error variance (L) leads to an item reliability = 0.95. These values cover the range of indicator reliability studied in typical mediation analyses simulation studies (e.g., [Fritz et al., 2016](#)). Note that low (high) measurement error in the indicators implies high (low) correlation, similarity among the indicators measuring the corresponding latent variable.

(c) Sample size: The results of the illustrative simulation in the paper in Figure 6 already suggest consistency of Bayes factor in our procedure. Here we examine 3 different sample sizes to see how the amount of evidence provided by Bayes factor changes with sample size at different specifications. We have 3 different sample sizes in the simulation: small samples (N=50), average samples (N=350), and large samples (N=1000). The average sample size is in line with the average sample size of the published studies we re-analyzed (N=319), and is meant to reflect the typical sample size used in the mediation literature.

(e) Number of mediator and outcome variables' indicators: We vary the number of indicators used for the latent mediator and outcome variables. As discussed in the paper and in Online Appendix B, we need at least two indicators for every latent variable to achieve identification. Therefore, the lower bound for the indicator variables in the study is $M_{ind} = 2, Y_{ind} = 2$. The upper bound is set to $M_{ind} = Y_{ind} = 4$. Again,

²⁴For this specification we only consider datasets with a zero direct effect. Directional misspecification induces a spurious direct effect, which leads to violation of conditional independence.

we opted for this upper bound based on the average indicator numbers of the published studies we re-analyzed ($E(M_{ind}) = 4.2, E(Y_{ind}) = 4.8$).

(f) Non-normality of measurement error distribution: Following [Tofghi and Kelley \(2020\)](#) and [Finch et al. \(1997\)](#), in addition to normally distributed measurement errors, we simulate samples with moderately non-normal condition (skewness=2, kurtosis=7), and severely non-normal condition (skewness=3, kurtosis=21) to examine the effect of error non-normality on the accuracy of our results.

The results of the simulation study show that the procedure is robust to misspecifications, as well as different sample characteristics. That is, the Bayes factor will only show evidence in favor of conditional independence in samples, where there is full causal mediation at work (and the mediation model estimated is correctly specified). These samples are associated with cells in Table E.1 where $\beta_3 = 0$. Note that small samples (N=50) do not convey enough data based information to support model choice. Bayes factor values for these samples show insufficient evidence for decisions about the model. To the contrary, samples with size N=1000 yield strong evidence reflected in their Bayes factor values. Moreover, Bayes factors of samples with size N=350, fall well within the region of substantive evidence. This suggests that the average sample collected in the literature provides enough information for the procedure to find evidence in favor of/against conditional independence that our sample of re-analyzed studies is representative). However we generally recommend collecting large samples when the goal is model choice, discovery.

Table E.4 and E.7 contain Bayes factors for models with M-Y unobserved confounder, and directionally misspecified mediation models. Both the unobserved confounder and misspecification of the causal direction between M and Y lead to conditional *dependence* in the data, which is reflected in the Bayes factor. These results too are robust to the different effect sizes, measurement error values and distributions investigated here, as well as to the number of indicators for latent variables. Larger sample sizes in these two cases lead to stronger evidence *against* conditional independence, and hence Bayes factors closer to 0.

Tables E.2, E.5, and E.8 show that Bayes factor detects presence and absence of conditional independence in datasets with high accuracy. The exceptions are small datasets (N=50) with full mediation. As discussed above, in these cases Bayes factor

values are in the interval $\{0.32, 3.2\}$, reflecting insufficient data-based evidence for model choice.

Finally, percentages of datasets with significant indirect effects in tables E.3, E.6, and E.9 show the indeterminacy of the results based on the analysis of indirect effects alone. In all the specifications, indirect effect estimates are significant with a high probability in larger datasets ($N=350$, and $N=1000$), and with a slightly lower probability in small datasets ($N=50$). Therefore, distinguishing the causal mediation model from models with no mediation, based on indirect effect estimates, is impossible.

1.1 Correctly specified mediation model (Bayes factors values)

U_{m^*}, U_{p^*} distribution		Normal						Moderately non-normal						Severely non-normal					
		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2	
Multiple indicators		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2	
$\sigma_{m^*}, \sigma_{p^*}, \beta_3$	β_1, β_2	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
LL	0	0.36	2.2	7.5	12.7	0.5	7.1	13.3	2.8	6.8	12.4	2.9	7.7	12.9	2.4	7.6	12.7	2.6	7.2
	0.48	2.1	6.8	12	0.5	7.2	12.3	2.7	7.3	12.5	2.6	7.4	12.3	2.5	7.5	12.3	2.3	6.8	12.4
	0.62	2.2	7	12.1	0.5	7	11.4	2.5	6.9	12.6	2.5	7.1	12.2	2.2	6.8	12	2.2	6.4	11.8
	0.36	0.36	2	0	1.5	0.1	0	0.8	0	0	0.7	0	0	0.7	0	0.6	0	0.7	0.1
	0.48	2.1	0.1	0	1.9	0.1	0	0.7	0.1	0	0.6	0	0	2	0.1	0	0.6	0	0.6
	0.62	2.1	0.1	0	2	0.1	0	2.3	0.1	0	0.7	0	0	2.2	0	2.2	0.1	0	1.9
LH	0	0.36	0.2	0	0.5	0	0	0.2	0	0	0.2	0	0	0.2	0	0.1	0	0.2	0
	0.48	0.4	0	0	1.7	0	0	0.2	0	0	0.4	0	0	0.2	0	0.1	0	0.4	0
	0.62	0.6	0	0	0.5	0	0	0.2	0	0	0.2	0	0	0.2	0	0.1	0	0.4	0
	0.36	2.2	6.9	11.4	2.6	6.4	11.8	2.4	6.7	12	2.8	7.2	12.7	2.2	6.9	11.9	2.2	6.9	11.3
	0.48	2.3	6.7	11.5	2.5	6.7	12.5	2.4	6.9	11.3	2.7	7.3	12.4	2.3	7.4	12.4	2.3	7	12
	0.62	2.2	5.9	11	2.4	6.9	11.9	2.3	6.2	11.5	2.3	6.9	11	2.3	7.1	12	2.2	6.6	11
HL	0	0.36	2.1	0	1.2	0.1	0	2.2	0.1	0	1.9	0.2	0	2.3	0.1	0	2	0.1	0
	0.48	2.1	0.2	0	2.1	0.1	0	2.1	0.1	0	2.6	0.1	0	2.4	0.1	0	2.2	0.1	0
	0.62	0.6	0.1	0	2	0.2	0	2.2	0.2	0	2.3	0.2	0	2.4	0.1	0	2	0.2	0
	0.36	0.5	0	0	0.4	0	0	0.3	0	0	0.2	0	0	0.4	0	0	0.6	0	0
	0.48	0.4	0	0	0.6	0	0	0.4	0	0	0.1	0	0	0.4	0	0	0.3	0	0
	0.62	0.5	0	0	0.6	0	0	0.4	0	0	0.5	0	0	0.4	0	0	0.5	0	0
HH	0	0.36	2	7.2	12.2	1	6.2	11.4	2.4	7.5	12.4	2.4	7.1	12.8	2.2	7.3	12.8	2.2	7
	0.48	1.9	6.5	11.9	1	6.3	10	2.2	7.1	11.8	2.4	6.7	12.1	2	6.3	11.7	2.3	6.6	11.5
	0.62	1.8	5.9	11.2	0.8	5.6	10.9	2.3	6.7	11.8	2.4	6.2	11.1	1.7	6.2	10.9	2.4	6.7	10.9
	0.36	2	0.1	0	1.4	0.1	0	0.7	0	0	0.7	0	0	1.6	0	0	0.6	0	0
	0.48	1.8	0.1	0	1.4	0.4	0	2.2	0.1	0	1.9	0.1	0	1.9	0.1	0	2	0.1	0
	0.62	1.9	0.2	0	1.5	0.6	0	1.9	0.1	0	1.9	0	0	2	0.1	0	2.1	0.1	0
HH	0	0.36	0.4	0	0	1.4	0	0	0.3	0	0	0.2	0	0	0.3	0	0	0.4	0
	0.48	0.4	0	0	1.3	0	0	0.3	0	0	0.3	0	0	0.4	0	0	0.5	0	0
	0.62	0.6	0	0	1.5	0	0	0.2	0	0	0.3	0	0	0.4	0	0	0.5	0	0
	0.36	2.2	6.4	11.5	2.3	7.1	12	2.2	6.8	11.2	2.4	7.5	12.1	2.3	6.3	11.5	2	6.9	11.7
	0.48	2.2	6.5	10.3	2.1	6.4	11.1	2.5	6.7	11.2	2.2	6.8	11.9	2.2	6.4	11	2.5	6.5	11.5
	0.62	2.2	6.2	10.1	2.2	6.4	10.7	2.3	6.4	10.8	2.4	6.8	11.3	2.1	6.1	10.8	2	6.6	11.7
HH	0	0.36	1.7	0.1	0	1.8	0.1	0	2.2	0.1	0	2.2	0	2	0.1	0	2.2	0.1	0
	0.48	1.9	0.1	0	2.1	0.1	0	2	0.1	0	2.5	0.1	0	1.8	0.1	0	2.3	0.1	0
	0.62	2	0.1	0	1.8	0.1	0	2.1	0.1	0	2.2	0.1	0	2.2	0.1	0	2.2	0.1	0
	0.36	0.3	0	0	0.4	0	0	0.3	0	0	0.4	0	0	0.4	0	0	0.4	0	0
	0.48	0.4	0	0	0.6	0	0	0.5	0	0	0.2	0	0	0.5	0	0	0.5	0	0
	0.62	0.5	0	0	0.7	0	0	0.5	0	0	0.5	0	0	0.4	0	0	0.5	0	0

Table E.1: The modes of the distribution of Bayes factors computed for datasets with mediation. Different conditions in the data are changed across the cells to examine the robustness of the Bayes factor.

1.2 Correctly specified mediation model (percentage of correct Bayes factor values)

1.3 Correctly specified mediation model (percentage of significant indirect effects)

U_{m*}, U_{p*} distribution		Normal						Moderately non-normal						Severely non-normal											
Multiple indicators		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4	
$\sigma_{m*}, \sigma_{p*}, \beta_3$	β_1, β_2	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
LL	0	0.66	1	1	0.71	1	1	0.8	1	1	0.66	1	1	0.61	1	1	0.72	1	1	0.52	1	1	0.68	1	1
		0.48	0.74	1	0.85	1	1	0.78	1	1	0.76	1	1	0.74	1	1	0.86	1	1	0.77	1	1	0.8	1	1
		0.62	0.75	1	0.87	1	1	0.92	1	1	0.86	1	1	0.85	1	1	0.94	1	1	0.91	1	1	0.86	1	1
	0.36	0.36	0.64	1	0.79	0.99	1	0.74	1	1	0.62	1	1	0.65	1	1	0.76	1	1	0.58	1	1	0.71	1	1
	0.48	0.69	1	0.86	1	1	0.86	1	1	0.76	1	1	0.8	1	1	0.84	1	1	0.7	1	1	0.71	1	1	
	0.62	0.84	1	0.89	1	1	0.91	1	1	0.82	1	1	0.83	1	1	0.89	1	1	0.85	1	1	0.89	1	1	
	0.62	0.36	0.67	1	0.72	1	1	0.74	1	1	0.68	1	1	0.62	1	1	0.75	1	1	0.63	1	1	0.62	1	1
	0.48	0.74	1	0.79	1	1	0.81	1	1	0.79	1	1	0.74	1	1	0.86	1	1	0.66	1	1	0.81	1	1	
LH		0.62	0.87	1	0.87	1	1	0.91	1	1	0.85	1	1	0.84	1	1	0.91	1	1	0.84	1	1	0.89	1	1
	0	0.36	0.58	1	0.68	1	1	0.63	1	1	0.62	1	1	0.56	1	1	0.66	1	1	0.52	1	1	0.65	1	1
		0.48	0.67	1	0.8	1	1	0.76	1	1	0.74	1	1	0.73	1	1	0.75	1	1	0.72	1	1	0.76	1	1
	0.62	0.82	1	0.84	1	1	0.92	1	1	0.79	1	1	0.82	1	1	0.89	1	1	0.82	1	1	0.88	1	1	
	0.36	0.36	0.68	1	0.72	1	1	0.74	1	1	0.58	1	1	0.61	1	1	0.68	1	1	0.5	1	1	0.7	1	1
	0.48	0.67	1	0.82	1	1	0.85	1	1	0.75	1	1	0.84	1	1	0.84	1	1	0.79	1	1	0.8	1	1	
	0.62	0.88	1	0.89	1	1	0.9	1	1	0.84	1	1	0.83	1	1	0.88	1	1	0.82	1	1	0.87	1	1	
	0.36	0.6	1	0.7	1	1	0.71	1	1	0.54	1	1	0.64	1	1	0.79	1	1	0.67	1	1	0.64	1	1	
	0.48	0.62	1	0.84	1	1	0.77	1	1	0.77	1	1	0.67	1	1	0.81	1	1	0.7	1	1	0.81	1	1	
	0.62	0.83	1	0.95	1	1	0.86	1	1	0.87	1	1	0.93	1	1	0.88	1	1	0.88	1	1	0.9	1	1	
HL	0	0.36	0.48	1	0.52	1	1	0.51	1	1	0.52	0.99	1	0.54	1	1	0.59	1	1	0.49	1	1	0.52	1	1
		0.48	0.59	1	0.72	1	1	0.71	1	1	0.62	1	1	0.61	1	1	0.7	1	1	0.71	1	1	0.72	1	1
		0.62	0.7	1	0.77	1	1	0.73	1	1	0.74	1	1	0.69	1	1	0.84	1	1	0.75	1	1	0.83	1	1
	0.36	0.36	0.52	1	0.54	1	1	0.61	1	0.99	0.58	1	1	0.63	1	1	0.69	1	1	0.59	1	1	0.63	1	1
	0.48	0.63	1	0.68	1	1	0.72	1	1	0.7	1	1	0.74	1	1	0.72	1	1	0.69	1	1	0.73	1	1	
	0.62	0.78	1	0.78	1	1	0.85	1	1	0.76	1	1	0.83	1	1	0.85	1	1	0.81	1	1	0.81	1	1	
	0.36	0.47	1	0.66	1	1	0.53	1	1	0.56	1	1	0.66	1	1	0.72	1	1	0.59	1	1	0.64	1	1	
	0.48	0.66	1	0.67	1	1	0.74	1	1	0.74	1	1	0.75	1	1	0.78	1	1	0.61	1	1	0.67	1	1	
	0.62	0.81	1	0.71	1	1	0.82	1	1	0.85	1	1	0.88	1	1	0.84	1	1	0.81	1	1	0.83	1	1	
HH	0	0.36	0.57	1	0.56	1	1	0.5	1	1	0.51	1	1	0.54	1	1	0.56	1	1	0.45	1	1	0.62	1	1
		0.48	0.67	1	0.63	1	1	0.76	1	1	0.64	1	1	0.66	1	1	0.67	1	1	0.58	1	1	0.76	1	1
		0.62	0.64	1	0.73	1	1	0.75	1	1	0.71	1	1	0.75	1	1	0.79	1	1	0.73	1	1	0.81	1	1
	0.36	0.36	0.48	0.99	0.45	1	1	0.54	1	1	0.56	0.99	1	0.43	1	1	0.6	1	1	0.39	1	1	0.55	1	1
	0.48	0.61	1	0.7	1	1	0.71	1	1	0.58	1	1	0.6	1	1	0.63	1	1	0.59	1	1	0.61	0.99	1	
	0.62	0.66	1	0.74	1	1	0.82	1	1	0.77	1	1	0.81	1	1	0.8	1	1	0.7	1	1	0.79	1	1	
	0.36	0.48	1	0.61	1	1	0.57	1	1	0.48	0.99	1	0.59	1	1	0.51	1	1	0.42	1	1	0.63	1	1	
	0.48	0.53	1	0.7	1	1	0.72	1	1	0.62	1	1	0.69	1	1	0.71	1	1	0.67	1	1	0.74	1	1	
	0.62	0.7	1	0.79	1	1	0.88	1	1	0.71	1	1	0.88	1	1	0.85	1	1	0.72	1	1	0.82	1	1	

Table E.3: Percentage of full and partial mediation datasets with a significant indirect effect estimate as indicated by NHST.

2.1 Model with unobserved confounder (Bayes factors values)

U_{m*}, U_{p*} distribution		Normal								Moderately non-normal								Severely non-normal							
		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4	
Multiple indicators		N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
$\sigma_{m*}, \sigma_{p*}, \beta_3$	β_1																								
LL	0	0.36	0.5	0	0	0.1	0	0	0.7	0	0	0.3	0	0	0.4	0	0	0.5	0	0	0.7	0	0	0.5	0
		0.48	0.4	0	0	0.1	0	0	0.4	0	0	0.2	0	0	0.3	0	0	0.3	0	0	0.5	0	0	0.2	0
		0.62	0.3	0	0	0.1	0	0	0.2	0	0	0.1	0	0	0.3	0	0	0.2	0	0	0.3	0	0	0.2	0
	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LH	0	0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0.36	0.5	0.1	0	0.9	0.1	0	0.7	0.1	0	0.6	0.1	0	1.9	0	0	2.2	0.1	0	2.1	0.1	0	0.4	0.1
		0.48	0.5	0.1	0	0.5	0	0	0.7	0.1	0	0.5	0	0	0.5	0	0	0.6	0	0	0.7	0.2	0	0.5	0.1
	0.36	0.62	0.6	0	0	0.4	0	0	0.5	0	0	0.4	0	0	0.3	0	0	0.4	0	0	0.6	0	0	0.4	0
	0.36	0.2	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0	0	0
		0.48	0.1	0	0	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0	0.1	0
		0.62	0.1	0	0	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0	0.1	0
	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HL	0	0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0.36	0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0.36	2	0	0	0.2	0	0	0.7	0	0	0.3	0	0	0.5	0.1	0	0.3	0	0	0.9	0	0	0.5	0
		0.48	0.5	0	0	0.2	0	0	2	0	0	0.3	0	0	0.3	0	0	0.4	0	0	0.4	0	0	0.5	0
		0.62	0.3	0	0	0.1	0	0	0.4	0	0	0.1	0	0	0.3	0	0	0.2	0	0	0.3	0	0	0.3	0
	0.36	0.36	0.1	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0
	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HH	0	0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		0.36	2.1	0.5	0	2.1	0.2	0	2	0.3	0	0.7	0.1	0	1.8	0.3	0	0.6	0.1	0	0.7	0.2	0	0.6	0.1
		0.48	1.9	0.2	0	0.5	0	0	1.8	0.2	0	0.6	0	0	0.8	0.1	0	0.6	0.1	0	0.6	0.1	0	0.7	0.1
	0.36	0.62	2	0	0	0.4	0	0	0.6	0	0	0.4	0	0	0.6	0	0	0.4	0	0	0.6	0	0	0.5	0
	0.36	0.36	0.3	0	0	0.1	0	0	0.1	0	0	0.2	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0	0
		0.48	0.1	0	0	0.1	0	0	0.2	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0
		0.62	0.1	0	0	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0	0.1	0	0	0.1	0	0	0.1	0
	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table E.4: The modes of the distribution of Bayes factors computed for datasets without mediation and with an M-Y unobserved confounder that induces an error correlation of $\rho_{MY} = -0.5$. Different conditions in the data are changed across the cells to examine the robustness of the Bayes factor.

2.2 Model with unobserved confounder (percentage of correct Bayes factor values)

U_{m*}, U_{p*} distribution		Normal								Moderately non-normal								Severely non-normal							
		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M4,Y4	
Multiple indicators		N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
LL	$\sigma_{m*}, \sigma_{p*}, \beta_3$	β_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
LH	0	0.36	1	0.95	0.99	1	0.92	0.99	1	0.93	1	0.97	1	1	0.96	1	1	0.86	0.91	1	1	0.94	1	1	0.99
	0.48	1	0.98	1	1	1	1	1	1	0.92	1	0.95	1	1	0.97	1	1	0.98	0.98	1	1	0.95	1	1	0.99
	0.62	1	0.98	1	1	1	1	1	1	0.99	1	1	1	1	1	1	1	0.99	1	1	0.98	1	1	1	0.97
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
HL	0	0.36	1	0.98	1	1	1	1	1	0.95	0.99	0.98	1	1	0.97	1	1	0.96	0.95	1	1	0.97	1	1	0.96
	0.48	1	0.96	1	1	1	1	1	1	0.97	1	1	1	1	1	1	1	0.98	0.98	1	1	0.99	0.98	1	1
	0.62	1	0.98	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.99	1	1	1	1	1	0.99	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
HH	0	0.36	1	0.8	0.99	1	0.93	1	1	0.81	0.98	0.94	0.99	1	0.93	1	1	0.9	0.88	1	1	0.92	1	1	0.88
	0.48	1	0.95	1	1	1	1	1	1	0.93	1	0.96	1	1	0.98	1	1	0.94	0.88	1	1	0.95	1	1	0.98
	0.62	1	0.96	1	1	1	1	1	1	0.98	1	0.97	1	1	0.94	1	1	0.97	0.95	1	1	0.97	1	1	0.98
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.36	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0.62	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Table E.5: Percentage of Bayes factors correctly identifying the absence of conditional independence ($\text{BF} < 3.2$).

2.3 Model with unobserved confounder (percentage of significant indirect effects)

U_{m*}, U_{p*} distribution		Normal								Moderately non-normal								Severely non-normal															
Multiple indicators		M2,Y2				M4,Y2				M2,Y4				M4,Y4				M2,Y2				M4,Y2				M2,Y4				M4,Y4			
$\sigma_{m*}, \sigma_{p*}, \beta_3$		N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50				
LL	0	0.63	1	1	0.63	1	1	0.76	1	1	0.54	1	1	0.71	1	1	0.74	1	1	0.54	1	1	0.66	1	1	0.69	1	1	0.69				
		0.48	0.67	1	0.84	1	1	0.84	1	1	0.65	1	1	0.88	1	1	0.8	1	1	0.72	1	1	0.72	1	1	0.81	1	1	0.81				
	0.36	0.62	0.77	1	0.83	1	1	0.92	1	1	0.71	1	1	0.95	1	1	0.95	1	1	0.75	1	1	0.75	1	1	0.85	1	1	0.85				
		0.36	0.55	1	0.75	1	1	0.68	1	1	0.51	1	1	0.68	1	1	0.71	1	1	0.49	1	1	0.68	1	1	0.74	1	1	0.74				
	0.48	0.71	1	0.8	1	1	0.8	1	1	0.67	1	1	0.82	1	1	0.76	1	0.82	1	0.73	1	1	0.82	1	1	0.83	1	1	0.83				
	0.62	0.73	1	0.86	1	1	0.87	1	1	0.75	1	1	0.93	1	1	0.89	1	0.93	1	0.82	1	1	0.84	1	1	0.91	1	1	0.91				
	0.36	0.47	1	0.64	1	1	0.71	1	1	0.48	1	1	0.64	1	1	0.65	1	0.78	1	0.58	1	1	0.63	1	1	0.69	1	1	0.69				
		0.48	0.7	0.75	1	1	0.76	1	1	0.69	1	1	0.63	1	1	0.75	1	0.8	1	0.71	1	1	0.81	1	1	0.83	1	1	0.83				
LH	0	0.62	0.81	1	0.87	1	1	0.85	1	1	0.82	1	1	0.89	1	1	0.88	1	0.88	1	0.74	1	1	0.84	1	1	0.86	1	1	0.86			
		0.36	0.64	1	0.64	1	1	0.73	1	1	0.5	1	1	0.64	1	1	0.64	1	0.71	1	0.51	1	1	0.66	1	1	0.67	1	1	0.67			
	0.48	0.59	1	0.74	1	1	0.69	1	1	0.67	1	1	0.73	1	1	0.69	1	0.72	1	0.68	1	1	0.68	1	1	0.82	1	1	0.82				
		0.62	0.76	1	0.76	1	1	0.82	1	1	0.7	1	1	0.8	1	1	0.74	1	0.81	1	0.78	1	1	0.81	1	1	0.81	1	1	0.81			
	0.36	0.36	0.49	1	0.55	1	1	0.67	1	1	0.51	1	1	0.74	1	1	0.77	1	0.77	1	0.58	1	1	0.66	1	1	0.67	1	1	0.67			
	0.48	0.66	1	0.73	1	1	0.78	1	1	0.62	1	1	0.84	1	1	0.73	1	0.78	1	0.63	1	1	0.74	1	1	0.78	1	1	0.78				
	0.62	0.67	1	0.84	1	1	0.85	1	1	0.72	1	1	0.85	1	1	0.75	1	0.84	1	0.73	1	1	0.77	1	1	0.8	1	1	0.8				
		0.36	0.54	1	0.6	1	1	0.68	1	1	0.49	0.99	1	0.57	1	1	0.57	1	0.73	1	0.49	1	1	0.65	1	1	0.65	1	1	0.65			
HL	0	0.48	0.61	1	0.69	1	1	0.72	1	1	0.72	1	1	0.71	1	1	0.74	1	0.74	1	0.62	1	1	0.68	1	1	0.79	1	1	0.79			
		0.62	0.67	1	0.85	1	1	0.91	1	1	0.8	1	1	0.8	1	1	0.87	1	0.87	1	0.57	1	1	0.82	1	1	0.85	1	1	0.85			
	0.36	0.48	0.46	1	0.53	1	1	0.59	1	1	0.48	0.98	1	0.57	1	1	0.51	1	0.51	1	0.49	1	1	0.61	1	1	0.66	1	1	0.66			
		0.62	0.66	1	0.84	1	1	0.79	1	1	0.58	1	1	0.54	1	1	0.68	1	0.67	1	0.58	1	1	0.71	1	1	0.73	1	1	0.73			
	0.36	0.36	0.41	1	0.5	1	1	0.52	1	1	0.37	1	1	0.5	1	1	0.6	1	0.6	1	0.41	1	1	0.53	1	1	0.57	1	1	0.57			
	0.48	0.56	1	0.65	1	1	0.61	1	1	0.6	1	1	0.66	1	1	0.66	1	0.73	1	0.52	1	1	0.7	1	1	0.69	1	1	0.69				
	0.62	0.65	1	0.73	1	1	0.72	1	1	0.52	1	1	0.52	1	1	0.66	1	0.78	1	0.66	1	1	0.75	1	1	0.78	1	1	0.78				
		0.36	0.51	1	0.44	1	1	0.53	1	1	0.36	1	1	0.36	1	1	0.57	1	0.49	1	0.37	1	1	0.55	1	1	0.67	1	1	0.67			
HH	0	0.48	0.53	1	0.68	1	1	0.64	1	1	0.43	1	1	0.69	1	1	0.72	1	0.72	1	0.47	1	1	0.64	1	1	0.66	1	1	0.66			
		0.62	0.56	1	0.65	1	1	0.75	1	1	0.52	1	1	0.71	1	1	0.8	1	0.8	1	0.64	1	1	0.76	1	1	0.79	1	1	0.79			
	0.36	0.43	1	0.45	1	1	0.52	1	1	0.44	0.99	1	0.54	1	1	0.54	1	0.55	0.99	0.4	1	1	0.56	1	1	0.52	1	1	0.52				
		0.48	0.52	1	0.45	1	1	0.68	1	1	0.59	1	1	0.55	1	1	0.55	1	0.77	1	0.51	1	1	0.59	1	1	0.64	1	1	0.64			
	0.36	0.36	0.35	1	0.48	0.99	1	0.51	1	1	0.47	1	1	0.54	1	1	0.64	1	0.64	1	0.33	0.99	1	0.58	1	1	0.56	0.99	1	0.56			
	0.48	0.47	1	0.62	1	1	0.55	1	1	0.45	1	1	0.49	1	1	0.59	1	0.49	1	0.42	1	1	0.64	1	1	0.6	1	1	0.6				
	0.62	0.54	1	0.58	1	1	0.67	1	1	0.65	1	1	0.65	1	1	0.65	1	0.67	1	0.53	1	1	0.7	1	1	0.78	1	1	0.78				
		0.36	0.32	1	0.5	1	1	0.53	1	1	0.37	1	1	0.44	0.99	1	0.46	1	0.54	1	0.36	1	1	0.53	1	1	0.57	1	1	0.57			
	0.48	0.35	1	0.51	1	1	0.63	1	1	0.45	1	1	0.45	1	1	0.52	1	0.63	1	0.4	1	1	0.62	1	1	0.67	1	1	0.67				
		0.62	0.54	1	0.64	1	1	0.7	1	1	0.58	1	1	0.65	1	1	0.65	1	0.75	1	0.54	1	1	0.73	1	1	0.74	1	1	0.74			

Table E.6: Percentage of datasets with an M-Y unobserved confounder with a significant indirect effect estimate as indicated by NHST.

3.1 Model with misspecified direction (Bayes factor values)

U_{m*}, U_{g*} distribution		Normal						Moderately non-normal						Severely non-normal					
Multiple indicators		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M2,Y2		M2,Y4	
$\sigma_{m*}, \sigma_{g*}, \beta_3$	β_1, β_2	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
LL	0	0.36	0	0	0.4	0	0	0.3	0	0	0.4	0	0	0.3	0	0	0.4	0	0
	0.48	0.3	0	0	0.4	0	0	0.2	0	0	0.3	0	0	0.4	0	0	0.2	0	0
	0.62	0.5	0	0	0.2	0	0	0.2	0	0	0.4	0	0	0.3	0	0	0.2	0	0
LH	0	0.36	0	0	0.4	0	0	0.2	0	0	0.5	0	0	0.4	0	0	0.4	0	0
	0.48	0.2	0	0	0.2	0	0	0.3	0	0	0.4	0	0	1.1	0	0	0.3	0	0
	0.62	0.3	0	0	0.4	0	0	0.1	0	0	0.4	0	0	0.5	0	0	0.4	0	0
HL	0	0.36	0	0	0.4	0	0	0.5	0	0	1.8	0	0	0.4	0	0	0.6	0	0
	0.48	0.5	0	0	0.5	0	0	0.5	0	0	0.5	0	0	0.6	0	0	0.5	0	0
	0.62	0.5	0	0	0.3	0	0	0.3	0	0	0.4	0	0	0.4	0	0	0.4	0	0
HH	0	0.36	0	0	0.4	0	0	0.4	0	0	0.5	0	0	0.5	0	0	0.4	0	0
	0.48	0.5	0	0	0.5	0	0	0.2	0	0	0.4	0	0	0.5	0	0	0.5	0	0
	0.62	0.6	0	0	0.4	0	0	0.3	0	0	0.5	0	0	0.7	0	0	0.4	0	0

Table E.7: The modes of the distribution of Bayes factors computed for datasets with mediation that are directionally misspecified. Different conditions in the data are changed across the cells to examine the robustness of the Bayes factor.

3.2 Model with misspecified direction (percentage of correct Bayes factor values)

U_{m*}, U_{g*} distribution		Normal						Moderately non-normal						Severely non-normal					
Multiple indicators		M2,Y2		M2,Y4		M4,Y2		M4,Y4		M2,Y2		M2,Y4		M4,Y2		M2,Y2		M2,Y4	
$\sigma_{m*}, \sigma_{g*}, \beta_3$	β_1, β_2	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000	N50	N350	N1000
LL	0	0.36	1	1	1	1	1	0.98	0.99	1	1	1	0.97	1	1	1	1	1	1
	0.48	1	1	1	1	1	1	1	1	1	1	1	0.99	1	1	1	1	1	1
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LH	0	0.36	1	1	1	1	1	1	1	1	1	1	0.97	1	1	1	1	1	1
	0.48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
HL	0	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0.48	1	1	1	1	1	1	1	1	0.98	1	1	1	1	1	1	1	1	1
	0.62	1	1	1	1	1	1	1	1	1	1	1	0.99	0.99	1	1	1	1	1
HH	0	0.36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0.48	1	1	1	1	1	1	1	1	0.99	1	1	1	1	1	1	1	1	1
	0.62	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table E.8: Percentage of Bayes factors correctly identifying the absence of conditional independence ($BF < 3.2$).

3.3 Model with misspecified direction (percentage of significant indirect effects)

U_{m^*}, U_{g^*} distribution		Normal												Moderately non-normal												Severely non-normal																											
		M2,Y2				M4,Y2				M2,Y4				M4,Y4				M2,Y2				M4,Y2				M2,Y4				M4,Y4				M2,Y2				M4,Y2				M2,Y4				M4,Y4							
Multiple indicators		$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2				$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2				$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2				$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2				$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2				$\sigma_{m^*}, \sigma_{g^*}, \beta_3$				β_1, β_2							
LL	0	0.36	0.28	0.86	1	0.34	0.9	1	0.27	0.82	1	0.32	0.86	1	0.26	0.81	1	0.34	0.89	1	0.29	0.85	0.99	0.47	0.84	1	0.27	0.87	1	0.31	0.85	1	0.31	0.85	1	0.27	0.87	1	0.31	0.84	1	0.31	0.84	1	0.27	0.87	1	0.31	0.84	1			
		0.48	0.42	0.97	1	0.51	0.96	1	0.33	0.93	1	0.45	0.96	1	0.46	0.98	1	0.59	0.98	1	0.47	0.96	1	0.46	1	1	0.39	0.95	1	0.44	0.98	1	0.44	0.98	1	0.39	0.95	1	0.44	0.98	1	0.44	0.98	1	0.39	0.95	1	0.44	0.98	1	0.39	0.95	1
		0.62	0.58	0.99	1	0.5	1	1	0.56	1	1	0.61	1	1	0.63	1	1	0.6	1	1	0.4	0.99	1	0.64	0.99	1	0.54	1	1	0.59	1	1	0.54	1	1	0.56	0.99	1	0.63	1	1	0.59	1	1	0.63	1	1	0.56	0.99	1			
LH	0	0.36	0.23	0.74	0.99	0.23	0.84	0.99	0.24	0.73	0.98	0.25	0.8	1	0.31	0.79	1	0.31	0.81	1	0.31	0.81	1	0.23	0.85	1	0.25	0.83	0.99	0.22	0.79	0.99	0.22	0.79	0.99	0.25	0.76	1	0.24	0.91	1	0.24	0.91	1	0.25	0.76	1	0.24	0.91	1			
		0.48	0.31	0.93	1	0.36	0.93	1	0.33	0.96	1	0.38	0.95	1	0.38	0.91	1	0.44	0.95	1	0.33	0.96	1	0.4	0.95	1	0.3	0.94	1	0.46	0.96	1	0.46	0.96	1	0.39	0.94	1	0.46	0.96	1	0.46	0.96	1	0.39	0.94	1	0.46	0.96	1			
		0.62	0.41	0.99	1	0.47	1	1	0.43	0.99	1	0.52	0.98	1	0.46	1	1	0.48	0.99	1	0.56	0.97	1	0.49	1	1	0.49	1	1	0.49	0.99	1	0.49	0.99	1	0.49	0.99	1	0.49	0.99	1	0.49	0.99	1	0.49	0.99	1	0.49	0.99	1			
HL	0	0.36	0.26	0.83	1	0.3	0.88	1	0.21	0.84	1	0.34	0.91	1	0.22	0.86	1	0.26	0.81	1	0.23	0.84	1	0.34	0.91	1	0.35	0.81	0.99	0.38	0.83	0.99	0.24	0.79	0.99	0.31	0.84	1	0.24	0.79	0.99	0.31	0.84	1	0.24	0.79	0.99	0.31	0.84	1			
		0.48	0.3	0.98	1	0.39	0.96	1	0.42	0.95	1	0.48	0.93	1	0.39	0.93	1	0.43	0.98	1	0.41	0.94	1	0.41	0.95	1	0.4	0.95	1	0.44	0.99	1	0.44	0.99	1	0.41	0.96	1	0.51	0.96	1	0.51	0.96	1	0.41	0.96	1	0.51	0.96	1			
		0.62	0.52	0.98	1	0.65	1	1	0.55	1	1	0.61	1	1	0.52	0.99	1	0.55	1	1	0.49	1	1	0.52	0.97	1	0.5	1	1	0.64	1	1	0.49	1	1	0.53	1	1	0.49	1	1	0.53	1	1	0.49	1	1	0.53	1	1			
HH	0	0.36	0.18	0.76	1	0.25	0.75	0.99	0.24	0.72	0.99	0.29	0.79	1	0.24	0.77	1	0.24	0.77	1	0.25	0.76	1	0.29	0.73	1	0.17	0.69	1	0.24	0.74	0.99	0.27	0.78	1	0.29	0.75	1	0.27	0.78	1	0.29	0.75	1	0.27	0.78	1	0.29	0.75	1			
		0.48	0.31	0.88	1	0.35	0.97	1	0.29	0.87	1	0.42	0.95	1	0.31	0.91	1	0.35	0.91	1	0.36	0.88	1	0.38	0.96	1	0.26	0.9	1	0.37	0.99	1	0.28	0.92	1	0.41	0.94	1	0.28	0.92	1	0.41	0.94	1	0.28	0.92	1	0.41	0.94	1			
		0.62	0.31	1	1	0.47	1	1	0.39	1	1	0.54	1	1	0.48	0.98	1	0.55	1	1	0.52	1	1	0.51	0.98	1	0.5	1	1	0.55	1	1	0.48	1	1	0.59	0.99	1	0.48	1	1	0.59	0.99	1	0.48	1	1	0.59	0.99	1			

Table E.9: Percentage of datasets with misspecified M-Y direction with a significant indirect effect estimate as indicated by NHST.

F Summary of re-analysis of published studies

We applied the proposed approach to mediation studies published in *Journal of Marketing Research* between February 2014 and December 2018 and in *Psychological Science* between January 2015 and September 2018.

Out of 285 articles in *Journal of Marketing Research* in the mentioned time period, 87 articles contain at least one mediation analysis. Out of these, 25 articles contain at least one "simple mediation". We approached the authors of these 25 articles. The authors of 8 articles agreed to share their data with us.

Regarding papers published in *Psychological Science* we could rely on articles with open-data badge. In the mentioned time frame, 53 articles contain at least one mediation analysis. Out of these, 34 study "simple mediation", i.e., mediation with one putative mediator. After excluding studies with interaction terms, we applied the proposed approach to sixteen studies. Fourteen of these use composite measures derived from multiple indicators to measure the mediator and/or the dependent variable.

We compare our results to those from using [Preacher and Hayes](#)'s bootstrapping approach in Table F.1. While [Preacher and Hayes](#)'s bootstrapping approach and the proposed method sometimes agree, the proposed method offers three improvements: First, it measures the evidence for mediation through Bayes Factor and distinguishes data based evidence for this causal model (see e.g., the results for the study by [Shaddy and Fishbach \(2017\)](#) and [Schroeder et al. \(2017\)](#) in Table F.1) from insufficient statistical information about a potential direct effect (see e.g., the results for [Nook et al. \(2018\)](#) and [Yan and Muthukrishnan \(2014\)](#) in Table F.1). Second, it affirms full mediation at the causal theory level, when the implied conditional independence relationship is obfuscated by measurement error or discretization (see e.g., the results for studies by [Sussman and O'brien \(2016\)](#) and [Zaval et al. \(2015\)\(1a\)](#) in Table F.1). Third, in cases where even after accounting for potential measurement biases conditional independence is rejected, the proposed approach improves inference for indirect effect estimates, if one is willing to assume mediation as the causal model a priori (see e.g., the results for [Van Zant and Moore, 2015\(1a and 2\)](#), and [Dai et al., 2015\(4a and 4b\)](#) in Table F.1).

Study	Bootstrapping results		Proposed method results				
	Indirect effect bootstrapped CI	Direct effect bootstrapped CI	Model used	Indirect effect 95% PCI	Direct effect 95% PCI	Bayes Factor (BF_{01})	BF numerical std. err.
Barasch et al. (2016), 2	[-1.61, -0.24]	[-4.55, -0.68]	LVM	[-1.74, -0.34]	[-2.62, 0.09]	0.25	0.001
Barasch et al. (2016), 3	[-3.91, -0.58]	[-9.50, -0.16]	LVM	[-4.18, -0.77]	[-2.55, 1.11]	0.79	0.014
Cheng & Cryder (2018)	[-42.59, -17.97]	[-32.09, 22.57]	Simple	[-42.46, -18.71]	[-1.97, 1.93]	1	10^{-6}
Dai et al. (2015), 4a	[0.003, 0.22]	[0.14, 0.95]	LVM	[0.04, 0.24]	[0.09, 0.90]	0.24	0.002
Dai et al. (2015), 4b	[0.01, 0.29]	[0.33, 1.23]	LVM	[0.04, 0.32]	[0.32, 1.15]	0.01	0.0007
Dong & Zhong (2018)	[-0.26, -0.08]	[-0.24, 0.08]	Simple	[-0.25, -0.09]	[-0.23, 0.07]	7.59	0.0005
Kleiman et al. (2016)	[-0.08, 0]	[-0.47, -0.03]	LVM	[-0.1, 0]	[-0.5, -0.03]	0.71	0.01
Moon et al. (2018)	[0.01, 0.1]	[0.78, 1.16]	LVM	[0.02, 0.12]	[0.8, 1.2]	0.0002	10^{-19}
Munichor & LeBoeuf (2018)	[0.36, 1.38]	[-0.62, 0.47]	Simple	[0.36, 1.32]	[-0.51, 0.50]	3.9	0.044
Nishikawa et al. (2017)	[0.22, 0.87]	[-0.57, 0.32]	Simple	[0.21, 0.86]	[-0.57, 0.28]	3.54	0.15
Nook et al. (2018)	[-0.46, -0.17]	[-0.39, 0]	Simple	[-0.48, -0.16]	[-0.38, 0]	1.58	0.004
O'Brien & Roney (2017)	[-0.61, -0.11]	[-0.46, 0.18]	Simple	[-0.61, -0.1]	[-0.43, 0.16]	4.5	0.0004
Romeo et al. (2018)	[0.02, 0.11]	[0.01, 0.16]	Simple	[0.02, 0.12]	[-0.01, 0.16]	4.49	0.01
Schroeder & Epley (2017)	[0.17, 0.53]	[-0.15, 0.24]	Simple	[0.17, 0.52]	[-0.15, 0.23]	9.27	0.003
Sela & LeBoeuf. (2017)	[-0.34, -0.05]	[-1.83, -0.90]	LVM	[-0.41, -0.07]	[-1.71, -0.79]	.000003	10^{-7}
Shaddy & Fischbach. (2017)	[-0.11, 0]	[-0.29, 0.05]	Simple	[-0.11, 0]	[-0.29, 0.06]	4.86	0.003
Sussman & O'Brien (2016), 3	[0.12, 0.31]	[-0.02, 0.33]	LVM	[0.20, 0.47]	[-0.14, 0.24]	8.55	0.08
Sussman & O'Brien (2016), 5	[0.54, 1.22]	[-0.62, 0.37]	Simple	[0.55, 1.21]	[-0.59, 0.36]	3.6	0.001
Tonietto & Malkoc (2016), 4a	[-0.17, -0.03]	[-0.3, 0.03]	Simple	[-0.17, -0.03]	[-0.3, 0.03]	3.27	0.004
Tonietto & Malkoc (2016), 4b	[0.09, 0.31]	[-0.11, 0.07]	LVM	[0.09, 0.32]	[-0.13, 0.08]	17.94	0.2
Tworek & Cimpian (2016)	[-0.12, -0.03]	[-0.18, 0.05]	Simple	[-0.12, -0.03]	[-0.18, 0.05]	9.11	0.001
Van Zant & Moore (2015), 1a	[0.33, 0.6]	[-0.3, -0.07]	LVM	[0.34, 0.63]	[-0.31, -0.07]	0.16	0.02
Van Zant & Moore (2015), 2	[0.24, 0.64]	[0.05, 0.43]	LVM	[0.28, 1.22]	[0, 0.4]	1.43	0.045
Yan & Muthukrishnan (2014)	[-0.76, -0.06]	[-1.36, 0.08]	Simple	[-0.76, -0.05]	[-1.31, 0.07]	0.65	0.001
Zaval et al. (2015), 1a	[0.01, 0.09]	[-0.08, 0.44]	LVM	[0.06, 0.3]	[-0.2, 0.33]	6.5	0.072
Zaval et al. (2015), 1b	[0.01, 0.08]	[0.05, 0.43]	LVM	[0.05, 0.33]	[-0.01, 0.55]	1.11	0.065

Table F.1: Comparison of the results from applying the proposed method on published mediation studies to the results from the analysis using bootstrapping

G Estimation algorithms

G.I Bayesian estimation and Bayes factor of the simple mediation model

Bayesian estimation: The joint posterior of the simple mediation model (Equations 1 - 3 in the paper) is as follows:

$$p(\beta_{0M}, \beta_1, \beta_{0Y}, \beta_2, \beta_3, \sigma_M^2, \sigma_Y^2 | X, M, Y) \propto p(Y | M, X, \beta_{0Y}, \beta_2, \beta_3, \sigma_Y^2) p(\beta_{0Y}, \beta_2, \beta_3) p(\sigma_Y^2) \\ p(M | X, \beta_{0M}, \beta_1, \sigma_M^2) p(\beta_{0M}, \beta_1) p(\sigma_M^2) \quad (14)$$

Where σ_M^2 and σ_Y^2 are variances of error terms U_M and U_Y respectively.

We specify the priors as follows:

$$\beta_{0M} \sim N(0, 100), \quad \beta_{0Y} \sim N(0, 100) \\ \beta_1 \sim N(0, 100), \quad \beta_2 \sim N(0, 100), \quad \beta_3 \sim N(0, 1) \quad (15) \\ \sigma_M^2 \sim Inv\chi^2(\nu, \nu S), \quad \sigma_Y^2 \sim Inv\chi^2(\nu, \nu S)$$

Default subjective prior values for ν and S are $\nu = 1, S = 3$.

Estimation employs the following algorithm:

1. Initialize parameters.
2. Draw $\beta_M = \{\beta_{0M}, \beta_1\}$ and σ_M^2 , with a Gibbs step, from their conditional posterior using a conditionally conjugate Normal-Inverse Gamma prior (see [Rossi et al., 2012](#)). We set the diagonal element of the prior precision matrix associated with β_3 to 1.
3. Draw $\beta_Y = \{\beta_{0Y}, \beta_2, \beta_3\}$ and σ_Y^2 , with a Gibbs step, from their conditional posterior, again using a conditionally conjugate Normal-Inverse Gamma prior.

Every aspect of the posterior distribution of the product $\beta_1 \cdot \beta_2$ can be obtained by summarizing products of posterior draws of β_1 and β_2 .

Bayes factor computation: To compute the Bayes factor based on the Savage Dickey ratio (see the derivations in Online Appendix C), we need to evaluate the marginal posterior of β_3 at $\beta_3 = 0$. A straightforward and efficient Monte-Carlo estimate of the

marginal posterior is:

$$p(\beta_3|X, M, Y) \approx \frac{1}{R} \sum_{r=1}^R N\left(\beta_3 | (\mu_{\beta_3})^r, (\sigma_{\beta_3}^2)^r\right) \quad (16)$$

Where $(\mu_{\beta_3})^r$ and $(\sigma_{\beta_3}^2)^r$ are the mean and the variance indexing the conditional posterior $p(\beta_3|M, Y, \beta_{0Y}, \beta_2, \sigma_Y^2)$ at each iteration $r = 1, \dots, R$ of the sampler iterating between $p(\sigma_Y^2|Y, M, \beta_{0Y}, \beta_2, \beta_3)$ and $p(\beta_{0Y}, \beta_2, \beta_3|Y, M, \sigma_Y^2)$.

G.II Bayesian estimation and Bayes factor of the multiple indicator mediation model with categorical observables

The latent variable model with K categorical indicators for M and L categorical indicators for Y is specified as follows:

$$\begin{aligned} M &= \beta_{0M} + \beta_1 X + U_M \\ m_1^* &= M + U_{m_1^*} \\ \tilde{m}_1 &= \text{OrdProbit}(m_1^*, C_{m_1}) \\ m_2^* &= \lambda_{20} + M + U_{m_2^*} \\ \tilde{m}_2 &= \text{OrdProbit}(m_2^*, C_{m_2}) \\ &\dots \\ m_K^* &= \lambda_{K0} + M + U_{m_K^*} \\ \tilde{m}_K &= \text{OrdProbit}(m_K^*, C_{m_K}) \\ Y &= \beta_{0Y} + \beta_2 M + \beta_3 X + U_Y \\ y_1^* &= Y + U_{y_1^*} \\ \tilde{y}_1 &= \text{OrdProbit}(y_1^*, C_{y_1}) \\ y_2^* &= \tau_{20} + Y + U_{y_2^*} \\ \tilde{y}_2 &= \text{OrdProbit}(y_2^*, C_{y_2}) \end{aligned} \quad (17)$$

$$\begin{aligned}
& \dots \\
y_L^* &= \tau_{L0} + Y + U_{y_L^*} \\
\tilde{y}_L &= \text{OrdProbit}(y_L^*, C_{y_L})
\end{aligned}$$

Note that to identify the model fixing only one of the cutpoints C for each indicator to 0, the intercepts of the first measurement equations are set to zero, and all slopes in the measurement equations as well as all (residual) variances of latent M and Y are fixed to one.

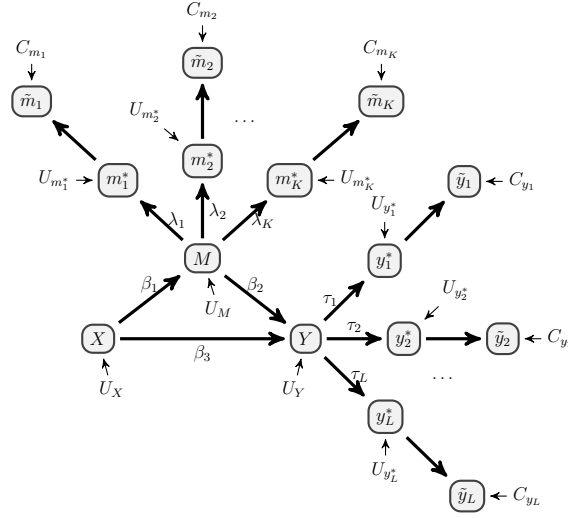


Figure G.1: Mediation model with multiple categorical indicators for M and Y

Bayesian estimation: The joint posterior of the multiple indicator mediation model is :

$$\begin{aligned}
& p(M, Y, \beta_{0M}, \beta_1, \beta_{0Y}, \beta_2, \beta_3, \sigma_M^2, \sigma_Y^2, \lambda_{20}, \dots, \lambda_{K0}, \sigma_{m_1^*}^2, \dots, \sigma_{m_K^*}^2, \\
& \tau_{20}, \dots, \tau_{L0}, \sigma_{y_1^*}^2, \dots, \sigma_{y_L^*}^2, C_{m_1}, C_{m_2}, \dots, C_{m_K}, C_{y_1}, C_{y_2}, \dots, C_{y_L}, m^*, y^* | X, \tilde{m}, \tilde{y}) \propto \\
& p(\tilde{y}_L | y_L^*, C_{y_L}) p(C_{y_L}) p(y_L^* | Y, \tau_{L0}, \sigma_{y_L^*}^2) p(\tau_{L0}) p(\sigma_{y_L^*}^2) \dots p(\tilde{y}_2 | y_2^*, C_{y_2}) p(C_{y_2}) p(y_2^* | Y, \tau_{20}, \sigma_{y_2^*}^2) p(\tau_{20}) p(\sigma_{y_2^*}^2) \\
& p(\tilde{y}_1 | y_1^*, C_{y_1}) p(C_{y_1}) p(y_1^* | Y, \sigma_{y_1^*}^2) p(\sigma_{y_1^*}^2) p(Y | M, X, \beta_{0Y}, \beta_2, \beta_3, \sigma_Y^2) p(\beta_{0Y}, \beta_2, \beta_3) p(\sigma_Y^2) \\
& p(\tilde{m}_K | m_K^*, C_{m_K}) p(C_{m_K}) p(m_K^* | M, \lambda_{K0}, \sigma_{m_K^*}^2) p(\lambda_{K0}) p(\sigma_{m_K^*}^2) \dots \\
& p(\tilde{m}_2 | m_2^*, C_{m_2}) p(C_{m_2}) p(m_2^* | M, \lambda_{20}, \sigma_{m_2^*}^2) p(\lambda_{20}) p(\sigma_{m_2^*}^2) p(\tilde{m}_1 | m_1^*, C_{m_1}) p(C_{m_1}) p(m_1^* | M, \sigma_{m_1^*}^2) p(\sigma_{m_1^*}^2) \\
& p(M | X, \beta_{0M}, \beta_1, \sigma_M^2) p(\beta_{0M}, \beta_1) p(\sigma_M^2)
\end{aligned} \tag{18}$$

We specify the priors as follows:

$$\begin{aligned}
\beta_{0M} &\sim N(0, 100), \quad \beta_{0Y} \sim N(0, 100) \\
\beta_1 &\sim N(0, 100), \quad \beta_2 \sim N(0, 100), \quad \beta_3 \sim N(0, 1) \\
\sigma_M^2 &\sim \text{Inv}\mathcal{X}^2(\nu, \nu S), \quad \sigma_Y^2 \sim \text{Inv}\mathcal{X}^2(\nu, \nu S) \\
\lambda_{20}, \dots, \lambda_{K0} &\sim N(0, 100) \\
\sigma_{m_1^*}^2, \dots, \sigma_{m_K^*}^2 &\sim \text{Inv}\mathcal{X}^2(\nu, \nu S) \\
\tau_{20}, \dots, \tau_{L0} &\sim N(0, 100) \\
\sigma_{y_1^*}^2, \dots, \sigma_{y_L^*}^2 &\sim \text{Inv}\mathcal{X}^2(\nu, \nu S)
\end{aligned} \tag{19}$$

Default subjective prior values for ν and S are $\nu = 1, S = 3$. In addition, we use the following prior for vector-valued variables

$$C^* \sim N(0, I),$$

where each observed indicator variable is related to such a vector-valued variable, $\{C_{m_i}^*\}_{i=1}^K$ and $\{C_{y_i}^*\}_{i=1}^L$. The elements in each vector-valued variable C^* are then exponentiated and cumulated to yield response indicator specific cut-off values relating unobserved continuous responses to observed discretized responses from rating scales.

We estimate the model with the following MCMC algorithm:

1. Initialize parameters.
2. Draw $\{C_{y_i}, y_i^*, \tau_{i0}, \sigma_{y_i^*}^2 | \tilde{y}_i, Y\}, i = 1, 2, \dots, L$, from an ordered probit model with draws of latent Y as regressor and regression coefficients fixed to 1.
3. Draw $Y | M, X, \beta_{0Y}, \beta_2, \beta_3, \{C_{y_i}\}_{i=1}^L, \{y_i^*\}_{i=1}^L, \{\tau_{i0}\}_{i=1}^L, \{\sigma_{y_i^*}^2\}_{i=1}^L$ from the following Normal distribution:

$$\begin{aligned}
p(Y | M, X, \beta_{0Y}, \beta_2, \beta_3, \{C_{y_i}\}_{i=1}^L, \{y_i^*\}_{i=1}^L, \{\tau_{i0}\}_{i=1}^L, \{\sigma_{y_i^*}^2\}_{i=1}^L) &\propto \\
p(\{y_i^*\}_{i=1}^L | Y, \{\sigma_{y_i^*}^2\}_{i=1}^L) &p(Y | M, X, \beta_{0Y}, \beta_2, \beta_3)
\end{aligned}$$

4. Draw $\{C_{m_i}, m_i^*, \lambda_{i0}, \sigma_{m_i^*}^2 | \tilde{m}_i, M\}, i = 1, 2, \dots, K$, from an ordered probit model with draws of latent M as regressor and regression coefficients fixed to 1.
5. Draw $M | Y, X, \beta_{0M}, \beta_{0Y}, \beta_1, \beta_2, \beta_3, \{C_{m_i}\}_{i=1}^K, \{m_i^*\}_{i=1}^K, \{\lambda_{i0}\}_{i=1}^K, \{\sigma_{m_i^*}^2\}_{i=1}^K$ from the following Normal distribution:

$$p(M|Y, X, \beta_{0M}, \beta_{0Y}, \beta_1, \beta_2, \beta_3, \{C_{m_i}\}_{i=1}^K, \{m_i^*\}_{i=1}^K, \{\lambda_{i0}\}_{i=1}^K, \{\sigma_{m_i}^2\}_{i=1}^K) \propto \\ p(\{m_i^*\}_{i=1}^K | M, \{\sigma_{m_i}^2\}_{i=1}^K, \{\lambda_{i0}\}_{i=1}^K) p(Y|M, X, \beta_{0Y}, \beta_2, \beta_3) p(M|X, \beta_{0M}, \beta_1)$$

6. Draw $\beta_{0M}, \beta_{0Y}, \beta_1, \beta_2, \beta_3 | X, M, Y$ using standard conjugate results.

We save the moments $\mu_{\beta_3}, \sigma_{\beta_3}^2$ indexing the full conditional posterior of β_3 in step 6 to construct an estimate of the marginal posterior of β_3 that we need to compute the Bayes factor as in Section G.I of this Appendix.

G.III Assessment of MCMC convergence

We show the convergence of MCMC chains for the simple model and the latent variable model with an illustrative simulation for each.

First we simulate data (N=500) from a simple mediation model as specified in Equations 2 & 3 with parameter values: $X \sim N(0, 1), \beta_1 = 0.3, \beta_2 = 0.5, \beta_3 = 0, Var(U_M) = 0.2, Var(U_Y) = 0.2$. Figure G.2 shows the traceplots of the MCMC chains (R=10000) of instances of parameters of the model (one coefficient and one error variance). The traceplots of both parameters show immediate convergence.

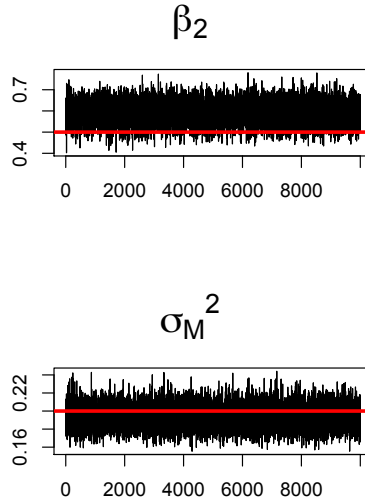


Figure G.2: Examples of the MCMC chains of the simple model posteriors of main coefficients (β_2) and error variances (σ_M^2). Red horizontal lines indicate the true value of the parameters.

Next we simulate data from a latent variable mediation model with four indicators for M and Y each. We simulate data (N=500) from the model specified in Equation 11, a latent variable model with 4 indicators for M and Y each, with parameter values: $X \sim N(0, 1),$

$\beta_1 = 1, \beta_2 = 1.5, \beta_3 = 0, \{\lambda_{10}, \lambda_{20}, \lambda_{30}, \lambda_{40}\} = \{0, 0.2, 0.3, 0.4\}, \{\tau_{10}, \tau_{20}, \tau_{30}, \tau_{40}\} = \{0, 0.1, 0.2, 0.3\}, C_{M_i} = C_{Y_i} = \{-100, 0, 1, 2, 3, 4, 5, 100\}, i = 1, \dots, 4, \text{Var}(U_{m_1}^*) = 0.5, \text{Var}(U_{m_2}^*) = 0.2, \text{Var}(U_{m_3}^*) = 0.5, \text{Var}(U_{m_4}^*) = 0.2, \text{Var}(U_{y_1}^*) = 0.3, \text{Var}(U_{y_2}^*) = 0.2, \text{Var}(U_{y_3}^*) = 0.3, \text{Var}(U_{y_4}^*) = 0.2$

A visual assessment of the MCMC chains of the instances of β coefficients, measurement error variances, indicator coefficients and cutoffs in Figure G.3 suggests that, the posteriors converge relatively fast, already after the first 500 draws, which are plotted separately from the rest of the draws. However in our applications we burn off the first 2000 draws to ensure that we are working with a fully converged posterior. We recommend researchers to run the MCMC for R=55000 iterations and throw the first 5000 draws as burnin. The short runtime of the MCMC algorithms (Simple model $\simeq 0.01s$, LVM $\simeq 28s$ for R=55000) allows for long MCMC runs, which guarantee the convergence of posteriors.

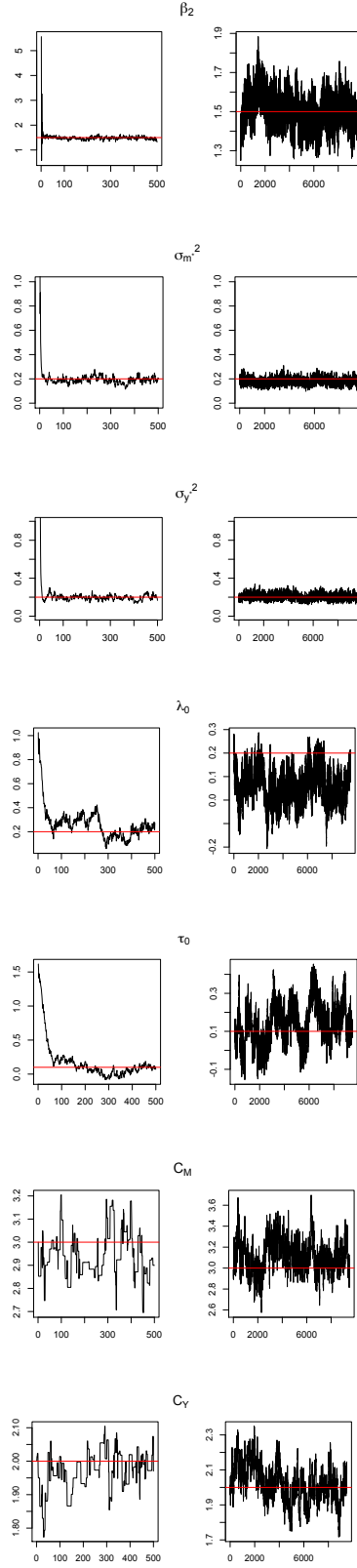


Figure G.3: Examples of the MCMC chains (burn-in periods plotted separately in the left panels) of the LVM model posteriors of main coefficients (β_2), indicator measurement error variances (σ_m^2, σ_y^2), intercepts of the measurement equations (λ_0, τ_0), and cutoff values (C_M, C_Y). Red horizontal lines indicate the true value of the parameters.