

ATTITUDES AND SOCIAL COGNITION

New Recommendations for Testing Indirect Effects in Mediational Models:
The Need to Report and Test Component PathsVincent Yzerbyt
Université catholique de LouvainDominique Muller and Cédric Batailler
Université Grenoble AlpesCharles M. Judd
University of Colorado

In light of current concerns with replicability and reporting false-positive effects in psychology, we examine Type I errors and power associated with 2 distinct approaches for the assessment of mediation, namely the component approach (testing individual parameter estimates in the model) and the index approach (testing a single mediational index). We conduct simulations that examine both approaches and show that the most commonly used tests under the index approach risk inflated Type I errors compared with the joint-significance test inspired by the component approach. We argue that the tendency to report only a single mediational index is worrisome for this reason and also because it is often accompanied by a failure to critically examine the individual causal paths underlying the mediational model. We recommend testing individual components of the indirect effect to argue for the presence of an indirect effect and then using other recommended procedures to calculate the size of that effect. Beyond simple mediation, we show that our conclusions also apply in cases of within-participant mediation and moderated mediation. We also provide a new R-package that allows for an easy implementation of our recommendations.

Keywords: indirect effects, mediation, joint-significance, bootstrap

Supplemental materials: <http://dx.doi.org/10.1037/pspa0000132.supp>

Techniques for the assessment of causal mediation have been the subject of extensive development for the last 30 years (Baron & Kenny, 1986; Judd & Kenny, 1981; Preacher, Rucker, & Hayes, 2007). The early literature recommended that mediation be demonstrated by examining and testing a set of individual parameter estimates within the overall model. We refer to this as the “component” approach that relies on joint-significance testing of multiple parameter estimates. Such practices were also recommended in more complex cases involving within-participant mediation (Judd, Kenny, & McClelland, 2001) and moderated mediation (Muller, Judd, & Yzerbyt, 2005). More recently, however, the literature has shunned this ap-

proach and has instead recommended an “index” approach of mediation (such as the PROCESS macro; Hayes, 2013), whereby trust in the underlying causal model rests on a single test of a mediational index. The purported advantages of this include a single statistical test rather than numerous ones, an increase in statistical power for mediation claims (a decrease in Type II errors; i.e., concluding there is no mediation effect when there is one in fact), and a single index that can be used to point to the plausibility of the underlying causal model. The downsides, from our perspective, are twofold. First, the reliance on a single index risks Type I errors (an increase in Type I error; i.e., concluding there is a mediation effect when there is none in fact).

Vincent Yzerbyt, Institute for Research in Psychological Sciences, Université catholique de Louvain; Dominique Muller and Cédric Batailler, Department of Psychology, Université Grenoble Alpes; Charles M. Judd, Department of Psychology, University of Colorado.

The authors are grateful to Sébastien Quittelier for his help in running the simulations on servers at the Université catholique de Louvain. They thank Arnold Ho, Nour Kteily, and Jacqueline Chen as well as Simone Dohle and Michael Siegrist for providing us with their data for illustration and inclusion in the JSmediation package. Portions of this work were presented by Dominique Muller at the 2014 General Meeting of the European Association of Social Psychology in Amsterdam. Vincent Yzerbyt, Dominique Muller, and Charles Judd are jointly responsible for the

ideas. Vincent Yzerbyt and Dominique Muller wrote the simulations. Vincent Yzerbyt ran them and compiled the results. Cédric Batailler conducted the journal surveys and created the JSmediation R package. Vincent Yzerbyt drafted the initial version of the article, and Charles Judd and Dominique Muller provided critical revisions. Dominique Muller and Cédric Batailler drafted the initial version of the supplementary material, and Vincent Yzerbyt and Charles Judd provided critical revisions. All authors approved the final version of the article for submission.

Correspondence concerning this article should be addressed to Vincent Yzerbyt, Institute for Research in Psychological Sciences, Université catholique de Louvain, Place du Cardinal Mercier, 10, B 1348 Louvain-la-Neuve, Belgium. E-mail: vincent.yzerbyt@uclouvain.be

Second, and importantly, such reliance may discourage researchers from ever examining the individual parameter estimates in the model or thinking critically about the model as a whole in light of these.

Our purpose is to revisit the distinction between the component approach (i.e., jointly testing individual parameter estimates in the model) and the index approach (i.e., computing and testing only a single mediational index, for instance a bias corrected bootstrap test). We present simulations that examine both Type I and Type II statistical error probabilities associated with the two approaches. We show that commonly used tests under the index approach risk inflated Type I errors, compared with the joint-significance test inspired by the component approach. Given the current concern with replicability and reporting false-positive effects in psychology,¹ we suggest that sole reliance on the index approach to mediation is worrisome, even while it does have small power benefits at least for the most biased of the index approach tests. Ultimately, we recommend testing and reporting individual components of the indirect effect to argue for mediation and then using other procedures to calculate the size and standard error of the indirect effect.

We devote most of our discussion to simple mediation but additionally demonstrate that our conclusions apply as well in cases of within-participant mediation and moderated mediation. We also present a new R package that allows for an easy implementation of our recommendations.

Simple Mediation: The Model and the Two Analytic Approaches

In psychology and related disciplines, a major research concern has been the search for intervening processes responsible for observed causal effects (Judd, Yzerbyt, & Muller, 2014). Specifically, the goal is to move beyond the demonstration of a relation between an independent variable and a dependent variable in order to provide evidence for some presumed underlying causal mechanism for that relation.

Illustrating Simple Mediation

In a recent study, Ho, Kteily, and Chen (2017, Study 3) hypothesized that telling African American participants that biracial individuals do versus do not experience discrimination would influence “hypodescent.” They defined “hypodescent” as the tendency to see biracial individuals as resembling more their low-status minority parent than their high-status one. They further argued that the relationship between the independent variable (the belief that biracials either do or do not experience discrimination) and hypodescent would be mediated by a sense of “linked fate” with biracial individuals. A total of 824 Black U.S. participants first read one of two articles that either made salient discrimination toward Black-White biracial individuals or did not (e.g., the article either claimed that such individuals experience discrimination in employment or claimed they did not). Participants then completed an eight-item linked-fate measure comprising such items as “Do you think what happens with Black people in this country will have something to do with what happens with Black-White biracials?,” “Black and Black-White biracials share a common fate,” and “Racial progress for Black people also means racial progress for Black-White individuals.” Finally, participants answered a three-item measure of “hypodescent” which began with the stem “If a Black American and a White American have a kid . . .”

followed by “would you think of the kid as relatively Black or relatively White?,” “would you consider the kid more Black or more White?,” and “how would categorize the kid?” For all three items, the scale ranged from 1 = *more Black* to 4 = *equally Black and White* to 7 = *more White* and was reverse-scored. Ho et al. (2017) conducted analyses of the resulting data that confirmed their hypotheses and we repeat these below following our exposition of methods for supporting claims of mediation.

Commonly used approaches for demonstrating a causal mediating process in data like these derive from recommendations of Baron and Kenny (1986; see also Judd & Kenny, 1981) that involve estimating three linear least-squares regression models (see Equations 1–3; b_{10} , b_{20} , and b_{30} are the intercepts in the three equations). The first examines whether the independent variable, X , affects the dependent variable, Y . The second examines the impact of the independent variable on the mediator, M . Finally, the third examines both X and M as simultaneous predictors of the dependent variable, Y . Assuming that the underlying assumptions about causal effects can be satisfied, mediation is claimed if the “total effect” of X on Y (c_{11}) is larger in absolute value than the “residual effect” of X on Y once M has been partialled out of both X and Y (c'_{31}).

$$Y = b_{10} + c_{11}X + e_1 \quad (1)$$

$$M = b_{20} + a_{21}X + e_2 \quad (2)$$

$$Y = b_{30} + c'_{31}X + b_{32}M + e_3 \quad (3)$$

In our example, estimating Model 1, participants indicated a higher belief in hypodescent in the high discrimination than in low discrimination condition, $c_{11} = 0.17$, $p < .04$. Second, estimating Model 2, discrimination condition also influenced the perception of linked fate, $a_{21} = 0.77$, $p < .001$. Finally, in Model 3 with both the perception of linked fate and condition (high vs. low discrimination) as predictors of judged hypodescent, condition was no longer predictive, $c'_{31} = 0.03$, $p = .74$, while the partial effect of linked fate was significant, $b_{32} = 0.19$, $p < .001$. Thus, the three linear models yielded results consistent with the authors’ predictions (see Figure 1).

If the models in Equations 1–3 are correctly specified, then there is an underlying equality to the mediational model, such that the total effect, c_{11} , is equal to the sum of the residual direct effect, c'_{31} , and the product of the other two effects in the model, a_{21} and b_{32} :

$$c_{11} = c'_{31} + a_{21}b_{32}$$

An algebraic re-expression of this yields the “fundamental equality” of mediation:

$$c_{11} - c'_{31} = a_{21}b_{32}$$

In other words, the difference between the total effect of X on Y and the residual direct effect once M is controlled (i.e., $c_{11} - c'_{31}$) must equal what is called the indirect effect: the product of the effect of X on M (a_{21}) and the partial effect of M on Y (b_{32}). In the data example, this algebraic equality is estimated as $0.17 - 0.03 = 0.77 \times 0.19$.

¹ An illustration of such concerns is the recent call in favor of using .005 rather than .05 as the critical threshold for claims of new discoveries (Benjamin et al., 2017; but see Krueger & Heck, 2017).

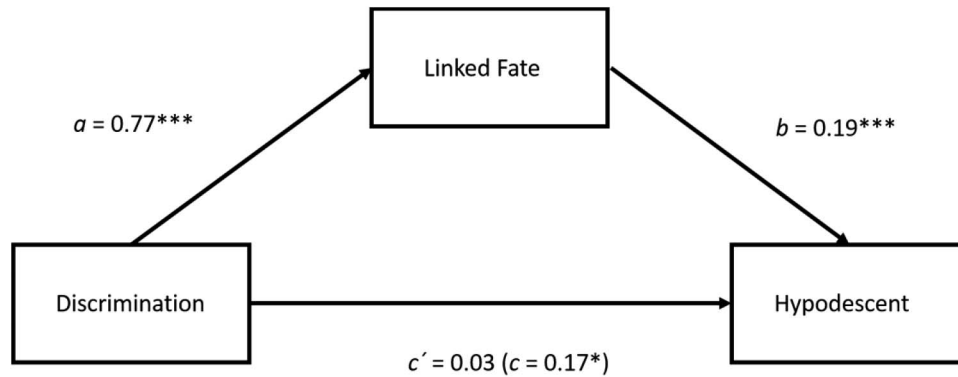


Figure 1. Mediation model from Ho et al. (2017, Study 3). Coefficients are unstandardized regression coefficients. The unstandardized regression coefficient representing the total relationship between condition and hypodescent is in parentheses. * $p < .05$. *** $p < .001$.

And the indirect effect is estimated as approximately 0.14.

The Two Approaches

Earlier we contrasted two general approaches to testing this indirect effect, calling one the component approach and the other the index approach. Both of these focus on the right side of the equality above, that is, $a_{21}b_{32}$, as the estimate of the indirect effect. The component approach proceeds by demonstrating that the two components of the indirect effect (a_{21} and b_{32}) are *both* significant. A test doing just that has been referred to as the “joint-significance” test (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) or the “causal steps” test (Biesanz, Falk, & Savalei, 2010). One compares both a_{21} and b_{32} with their respective standard errors, under normal distribution assumptions² (or one can use the two corresponding confidence intervals; Cumming, 2014). This amounts to the results one gets for a_{21} and b_{32} when testing them in Models 2 and 3 using ordinary least square regression. An indirect effect is claimed, according to the joint-significance test, only if *both* of these individual coefficients are simultaneously significant (or if neither of their confidence intervals includes 0). Going back to Ho et al.’s (2017) study, both a_{21} and b_{32} were highly significant, confirming the presence of an indirect effect.

In contrast to the component approach, the index approach to testing mediation is based on the assumption that there should only be *one* overall test of the indirect effect, rather than two separate tests of its different components. Thus, the index approach uses various methods to provide a statistical test of whether the $a_{21}b_{32}$ product as a whole, rather than its components individually, differs significantly from zero. The rationale provided for this, given in multiple publications (Hayes, 2013, 2015; for a recent example, see Montoya & Hayes, 2017, p. 7), is that multiple hypothesis tests are inherently problematic because Type I errors are inflated across multiple tests.

Before we examine the specific methods used to test the indirect effect under the index approach, it is important to say that this argument against the component approach (and its joint-significance test), appealing as it seems, is incorrect. The inflation of Type I errors by conducting multiple tests applies in cases where the overall null hypothesis is rejected if *any* of the tests yields a significant result. For instance, if one did two tests and

required that only one of them be significant, using for each $\alpha = .05$, the overall Type I error rate (known as familywise error rate) would be $1 - (1 - \alpha)^2$, which is .0975. However, with the joint-significance test, both a_{21} and b_{32} must be *simultaneously* significant for an indirect effect to be claimed, so this rationale against two tests is unwarranted.

The index approach tests the product $a_{21}b_{32}$ with one test of statistical significance, with the null hypothesis that the product equals zero. Many such tests exist in the literature. Early recommendations for testing this ab product derived its standard error from the pooled standard errors of the individual components, an approach frequently labeled the Sobel test (Baron & Kenny, 1986; Sobel, 1982). This test, however, has largely been abandoned due to violations of the normality assumption in testing the product of two coefficients (MacKinnon, Warsi, & Dwyer, 1995). Therefore, more robust methods are now the norm for testing the ab product.

The most widely used methods for testing the $a_{21}b_{32}$ product rely on resampling or bootstrapping procedures, in which one resamples observations with replacement from the original data, computes the $a_{21}b_{32}$ product in each new sample, and then examines the distribution of these products across many samples (for an overview, see Ong, 2014). In the present context, there are three different versions of this method. The first, the percentile bootstrap, computes the 95% percent confidence interval for the true value of $a_{21}b_{32}$, given all the resampled estimates. If the value of zero is outside of this interval, then one concludes that the estimated value in one’s data permits rejection of the null hypothesis that the indirect effect index equals zero. The second, the bias-corrected bootstrap, deals with the fact that the mean $a_{21}b_{32}$ product of the bootstrapped samples does not always equal the actual $a_{21}b_{32}$ product estimated in the data. This second method corrects for that. Finally, a third version, the accelerated bias-corrected bootstrap, adjusts for the fact that the variance of the $a_{21}b_{32}$ estimates across the bootstrapped samples varies. In recent years, the most widely used of the macros that was available

² If the OLS assumptions are violated for any of the paths considered in the joint-significance approach, one would need to consider alternative means of checking for significance, with bootstrap being one possible strategy.

(<http://www.processmacro.org/index.html>) relied on these bootstrapping procedures with the default being the bias-corrected bootstrap (but see the PROCESS 3.0 macro released early 2018).

Another method avoids, like bootstrapping, the distributional assumptions of traditional statistical inference. Known as the Monte Carlo test, it uses the a_{21} and b_{32} estimates and their standard errors (MacKinnon, Lockwood, & Williams, 2004). Assuming these come from two normal distributions, it independently samples individual values of each from those underlying distributions and then computes the product of the sampled values. This is repeated a very large number of times, generating again a confidence interval for the true $a_{21}b_{32}$ product value. A distinct advantage here is that the confidence interval can be generated without access to the raw data because only the a_{21} and b_{32} estimates and their standard errors are necessary (for an implementation in R, see Tofighi & MacKinnon, 2011 and the R package we provide with this article).

Comparative Performance of Methods

Several articles have evaluated these various tests in terms of their susceptibility to both Type I and Type II errors (Biesanz et al., 2010; Fritz & MacKinnon, 2007; Fritz, Taylor, & MacKinnon, 2012; Hayes & Scharkow, 2013; MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, & Williams, 2004). Most have focused on the four tests of the index approach (i.e., percentile bootstrap, bias-corrected bootstrap, accelerated bias-corrected bootstrap, and Monte Carlo) and have shown somewhat superior power for the bias-corrected bootstrap and the accelerated bias-corrected bootstrap than for the other index approach tests. As a result, all else being equal, this means that slightly fewer observations are required to demonstrate a significant indirect effect using these two tests. On the other hand, starting with MacKinnon, Lockwood, and Williams (2004), there are demonstrations of inflated Type I error rates for the two bias-corrected tests, mostly when the true value of either a_{21} or b_{32} (but not both) equals zero. As a result, Fritz, Taylor, and MacKinnon (2012) recommend the two bias-corrected tests if statistical power is the major concern, but the percentile bootstrap or Monte Carlo tests if Type I errors are more worrisome. Reservations regarding the use of the accelerated bias-corrected test also emerged from extensive simulations by Biesanz, Falk, and Savalei (2010). These authors actually conclude that the accelerated bias-corrected test should not be used due to unacceptably inflated Type I error rates.

Although the joint-significance test, based on the component approach to testing the indirect effect, was sometimes included in these studies, the lessons that could be drawn from comparison of it with the index approach tests were generally overlooked. First, in terms of statistical power, Fritz and MacKinnon's (2007) and Biesanz et al.'s (2010) simulations showed that the joint-significance test was at least as powerful as the percentile bootstrap (in fact, it was often more powerful), while these two tests were a bit less powerful than the two adjusted bootstrap tests. This pattern confirms our earlier comment that the component approach, relying on two tests of individual coefficients, is not inherently less powerful than an index test that relies on a single test. Additionally, in terms of Type I error, previous work rarely included or discussed the relative performance of the joint-significance test when the true value of either a_{21} or b_{32} (but not both) equals zero. One notable exception can be found in the work

by Biesanz et al. (2010) who examined several methods for testing indirect effects using a variety of data structures (i.e., normal and non-normal, complete, and incomplete data sets).

As we suggested earlier, recent difficulties in replicating effects in psychology underline the importance of holding Type I error rates at appropriate and known levels. Accordingly, comparisons of alternative testing procedures in terms of their statistical power should preferably be done only when it is known that those procedures do not yield inflated Type I error rates. The existing literature fails to provide definitive answers about the comparison between the component and the index approaches in terms of Type I error rates. Therefore, we conducted simulations to examine the performance of the various tests, focusing on the differences between the joint-significance test and the various tests of the ab index. In doing so, we partially replicate portions of the extensive simulations conducted by Biesanz et al. (2010). These authors' efforts have remained relatively unknown and it seems useful to examine the reliability of their conclusions.

Beyond the convergence of our results with those of Biesanz et al. (2010), our simulations provide focus and extensions that should be particularly compelling for social psychologists. First, we consider the bias-corrected bootstrap index approach, which was not included in the Biesanz et al. (2010) simulations, allowing us to compare the performance of this method with the joint-significance test. This is important because it is this index method that has been the default recommended option of a popular mediation package, that is, the PROCESS macro, until very recently and because our review of the literature (reported later) suggests that this approach is very widely used. The second very important extension is that we develop our argument well beyond the case of simple mediation, exploring index versus component approaches in situations that social psychologists frequently encounter, specifically within-subject mediation models as well as moderated mediation models.

In sum, and importantly, although our simulations end up reinforcing Biesanz et al.'s (2010) conclusions about the joint-significance test (what they call the causal steps approach), those conclusions were not the central focus of their article. Our article is focused on the distinction between what we call the component and index approaches, thus making salient for readers the importance of examining individual parameter estimates in mediation models. Ultimately, our message is that researchers should not simply report and test a single index and assume that doing so demonstrates and capture all there is to mediation. That message is not conveyed in the Biesanz et al. (2010) article, in spite of its many virtues.

Simple Mediation: The Simulations

The simulations we report are similar to those reported by others (e.g., Biesanz et al., 2010; Fritz & MacKinnon, 2007; Hayes & Scharkow, 2013).³ A simple mediation model was assumed, varying sample sizes and true parameter values for a_{21} and b_{32} . In light of samples typically used in psychology, the three sample sizes used were 50, 100, and 200. True values for a_{21} and b_{32} were set at .00, .14, .39, and .59, corresponding to zero, small, medium, and large effects (given the error variances used). Every combination

³ All simulations presented in this article were computed by means of dedicated programs using R version 3.1.0 (R Core Team, 2014). The programs are available from the authors upon request.

of the two values for a_{21} and b_{32} were used. The true value of c'_{31} was always set to zero. Values of X were sampled from a normal distribution with mean 0 and variance 1. Errors to both M and Y were similarly sampled. For each combination of true parameter values and sample sizes (48 different combinations), 10,000 samples were generated and mediational analyses conducted in each sample. In 21 of the 48 different combinations of sample size and parameter values, the null hypothesis of no mediation is correct (i.e., an indirect effect of zero because one or both of a_{21} and b_{32} equals zero). In the remaining 27 cases, there is in fact true mediation. Type I errors were examined in the first context (finding a significant mediation effect when none is present) and power was examined in the second context (finding a significant mediation effect when such an effect is present).

In each sample, we compared five different procedures to test for the presence of mediation. First, the joint-significance test was used, testing the a_{21} and b_{32} components individually. The other four tests were index approaches, testing the significance of the $a_{21}b_{32}$ product with the four methods discussed earlier: Monte Carlo, percentile bootstrap, bias-corrected bootstrap, and adjusted bias-corrected bootstrap.

For the joint-significance test, we computed two ordinary least-squares regression analyses (see Equations 2 and 3, above) for each sample and estimated coefficients a_{21} and b_{32} along with their respective standard errors. An indirect effect was declared significant when *both* coefficients a_{21} and b_{32} were significant.

In order to calculate the Monte Carlo confidence interval for each of the 10,000 samples, we sampled 1,000 independent random pairs of normal deviates with means a_{21} and b_{32} and their standard errors from the above regressions. The standard normal deviates of each pair were then multiplied to produce a distribution for each sample and the endpoints of the 95% confidence interval were calculated as the products defining the 2.5th and the 97.5th percentile values of the distribution. An indirect effect was declared significant for the given sample if the confidence interval failed to include zero.⁴

For the bootstrap methods, for each of the 10,000 samples, we generated 1,000 bootstrap samples of the same number of cases by sampling from the sample with replacement. Fritz and MacKinnon (2007) indicated that a higher number of bootstrap samples did not affect the proportion of Type I errors. Coefficients a_{21} and b_{32} , along with their product, $a_{21}b_{32}$, were calculated for each bootstrap sample. These products produce a distribution across the samples and the endpoints of the 95% confidence interval are calculated as the products defining the 2.5th and the 97.5th percentile values of that distribution. Again, an indirect effect was declared significant for the sample if the confidence interval failed to include zero. The bias-corrected and accelerated bias-corrected were based on the same initial bootstrap samples, but adjusted for bias using the bias adjustments proposed by Hayes (2013), MacKinnon (2008), and Preacher and Selig (2012).

Type I Errors

For 21 of the 48 simulations, at least one of the two true parameter values, a_{21} and b_{32} , was zero. As a result, their true product must be zero, corresponding to the absence of an indirect effect. Results for these are given in the top panel of Table 1. Values given are the proportion of samples (out of 10,000) yield-

ing a significant indirect effect (empirical Type I error rates), using alpha of .05. Values above .05 indicate that the error rate for a particular test is inflated. Regardless of sample size and parameter values, the prevalence of Type I errors for the joint-significance test never exceeds .05. This is not true for any of the four index methods. Whenever one of the parameter values was either moderate or large, and the other one was zero, both the bias-corrected and the adjusted bias-corrected bootstrap procedures yielded empirical Type I error rates that were consistently larger than .05. The other two index methods (Monte Carlo and percentile bootstrap) yielded more appropriate Type I error rates, but even these were frequently above the .05 value. In sum, only the joint-significance test showed no inflation in Type I error rates.

Power

In 28 of our simulations, there was a true indirect effect, because both of the true parameter values of a_{21} and b_{21} were nonzero. In these cases, the matrix at the bottom of Table 1 gives power values (proportion of samples in which a significant indirect effect was found) for each test. Unsurprisingly, higher power is found with larger sample sizes. Additionally, the methods can be grouped into two clusters, with the joint-significance test, the Monte Carlo sampling method, and the percentile bootstrap method on the one hand, and the two bias-corrected methods on the other. Power is consistently higher for the two bias-corrected methods than for the other three. Importantly, power for the joint-significance test is equal to that obtained for the Monte Carlo and percentile bootstrap approaches.

Type I Method Inconsistencies

In addition to these results, it is important to examine the number of samples in which pairs of methods yield conflicting results (Hayes & Scharkow, 2013), particularly in the case of Type I errors (i.e., one method finds a significant effect in the absence of a true effect while the second does not). Given our focus on the component approach and the index approach, we compared the joint-significance test to the Monte Carlo method and, more specifically, the two bootstrapping methods (percentile and bias-corrected) in terms of inconsistencies in those cases where no true indirect effect exists. Table 2 shows, for each pair of tests, the number of times (out of 10,000 simulated samples) one method yielded a significant effect (Type I error) and the other did not. The joint-significance test fares relatively better than the percentile bootstrap method, the latter leading to 1.67 times more unique Type I errors. The comparison with the bias-corrected bootstrap method is even more telling, with almost 14 times more unique Type I errors than the joint-significance test.

Implications and the Practice of Reporting Simple Mediation Analyses

The message emanating from the present simulations converges with the earlier simulation work conducted by Biesanz et al.

⁴ The distribution of the product method, discussed in MacKinnon, Fritz, Williams, and Lockwood (2007) can be seen as an alternative to the Monte Carlo approach. Here, we only relied on the latter as these methods are largely interchangeable and rarely produce different inferences (Hayes & Scharkow, 2013).

Table 1
Type I Errors (Top Panel) and Power (Bottom Panel) for the Simple Mediation Analysis as a Function of the Population Values of Coefficients a_{21} and b_{32} , Sample Size, and Method

	Type 1 error								
	a = 0			a = .14			a = .59		
	b = 0	b = .14	b = 0	b = .39	b = 0	b = .59	b = 0	b = 0	b = 0
<i>n</i> = 50									
JS	.0020	.0072	.0069	.0343	.0406	.0486	.0474		
MC	.0016	.0062	.0063	.0351	.0409	.0528	.0537		
PB	.0025	.0080	.0074	.0391	.0438	.0595	.0604		
BC	.0084	.0200	.0199	.0698	.0765	.0864	.0869		
ABC	.0111	.0255	.0293	.0812	.0868	.0945	.0943		
<i>n</i> = 100									
JS	.0033	.0148	.0146	.0478	.0469	.0509	.0514		
MC	.0023	.0112	.0110	.0485	.0479	.0568	.0577		
PB	.0029	.0121	.0114	.0541	.0500	.0626	.0609		
BC	.0073	.0314	.0295	.0817	.0792	.0779	.0758		
ABC	.0087	.0372	.0355	.0864	.0834	.0814	.0766		
<i>n</i> = 200									
JS	.0027	.0239	.0261	.0519	.0503	.0508	.0490		
MC	.0021	.0180	.0201	.0559	.0555	.0541	.0510		
PB	.0025	.0207	.0217	.0578	.0582	.0569	.0533		
BC	.0066	.0463	.0472	.0735	.0734	.0674	.0613		
ABC	.0079	.0520	.0501	.0751	.0773	.0687	.0620		
	Power								
	a = .14			a = .39			a = .59		
	b = .14	b = .39	b = .59	b = .14	b = .39	b = .59	b = .14	b = .39	b = .59
<i>n</i> = 50									
JS	.0248	.1150	.1534	.1160	.5523	.7257	.1502	.7298	.9448
MC	.0218	.1158	.1641	.1139	.5542	.7364	.1643	.7428	.9492
PB	.0240	.1177	.1729	.1172	.5543	.7314	.1636	.7304	.9395
BC	.0508	.1826	.2234	.1839	.6455	.7887	.2158	.7895	.9589
ABC	.0587	.1942	.2316	.1963	.6496	.7837	.2216	.7842	.9522
<i>n</i> = 100									
JS	.0759	.2629	.2851	.2783	.9311	.9625	.2793	.9604	.9996
MC	.0662	.2655	.2971	.2781	.9327	.9646	.2912	.9616	.9995
PB	.0657	.2695	.3053	.2848	.9302	.9643	.2949	.9595	.9995
BC	.1291	.3377	.3420	.3534	.9511	.9712	.3322	.9682	.9995
ABC	.1418	.3446	.3438	.3539	.9505	.9691	.3307	.9655	.9994
<i>n</i> = 200									
JS	.2545	.5073	.4991	.5070	.9987	.9996	.5033	.9994	1.0000
MC	.2290	.5174	.5072	.5156	.9988	.9996	.5123	.9994	1.0000
PB	.2290	.5190	.5105	.5130	.9986	.9996	.5120	.9996	1.0000
BC	.3402	.5636	.5316	.5563	.9990	.9996	.5299	.9996	1.0000
ABC	.3487	.5597	.5326	.5572	.9987	.9995	.5307	.9996	1.0000

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap; ABC = accelerated bias-corrected bootstrap. Bold numbers are Type I error rates significantly above .05, i.e., for which .05 falls below the lower limit of the 95% confidence interval.

(2010). These authors also note that, when it comes to null hypothesis testing, the joint-significance test (which they call the causal steps method) has the best balance of Type I error and statistical power. They also reached the same conclusions when they simulated non-normally distributed data and data sets with missing observations.

Biesanz et al. (2010) suggest that the accelerated bias-corrected method for testing the indirect effect should be discarded altogether because its Type I error rate is simply too high in some

cases (above .07, sometimes close to .10). They argue that its apparent power benefits are not worth the associated risk of alpha inflation when the null hypothesis is true. Our own simulations, which are entirely consistent with those of Biesanz et al. (2010), lead us to concur. Additionally, we have shown that the bias-corrected method, which was not considered by Biesanz et al. (2010), is equally problematic. This is an important finding insofar as the bias-corrected method has been the default method in the most popular mediation macro for a long time and until very

Table 2

Type I Errors Uniquely Due to One as Opposed to the Other Method as a Function of the Values of Coefficients a_{21} and b_{32} , Pair of Methods, and Sample Size in the Context of a Simple Mediation Analysis

Comparison	n	$a = 0$	$a = 0$	$a = .14$	$a = 0$	$a = .39$	$a = 0$	$a = .59$	Sub-Mean	Mean
		$b = 0$	$b = .14$	$b = 0$	$b = .39$	$b = 0$	$b = .59$			
JS = 1 & MC = 0	50	5	14	14	40	30	29	25	22.43	29.76
	100	14	46	40	40	42	18	14	30.57	
	200	7	67	68	27	30	26	29	36.29	
JS = 0 & MC = 1	50	1	4	8	48	82	71	88	43.14	40.38
	100	4	10	4	47	52	77	77	38.71	
	200	1	8	8	68	82	59	49	39.29	
JS = 1 & PB = 0	50	6	32	20	72	91	83	70	53.43	52.62
	100	12	53	57	73	77	51	64	55.29	
	200	9	66	72	52	40	51	54	49.14	
JS = 0 & PB = 1	50	11	40	25	120	123	192	200	101.57	88.05
	100	8	26	25	136	108	168	159	90.00	
	200	7	34	28	111	119	112	97	72.57	
JS = 1 & BC = 0	50	1	5	6	20	22	26	25	15.00	16.81
	100	2	6	13	20	17	26	34	16.86	
	200	1	10	12	25	14	29	39	18.57	
JS = 0 & BC = 1	50	65	133	136	375	381	404	420	273.43	233.48
	100	42	172	162	359	340	296	278	235.57	
	200	40	234	223	241	245	195	162	191.43	

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap. A 1 means that the method led to the decision of the presence of an indirect effect and a 0 means that the method led to the decision of an absence of indirect effect. For instance, when $a = 0$ and $b = 0$ and JS = 1 and MC = 0 and $n = 50$, 5 means that on five occasions (out of 10,000) the joint-significance test led to conclude in favor of an indirect effect when the Monte Carlo sampling method did not.

recently (Version 2.16 still relies on this method as its default method and it has been removed only in Version 3.00 which came out in early 2018: <http://www.processmacro.org/index.html>). Given the prevalence of this method we suspect that a substantial number of reported tests of mediation in the literature may be problematic.

The consistency of Biesanz et al.'s (2010) findings with our own leads us to believe that the most sensible way to approach an indirect effect is to start by examining the individual coefficients and establish that both of them are indeed significant. At the same time, and despite the obvious merits of the component approach advocated here, we acknowledge the fact that sole reliance on the joint-significance test does not provide researchers with a single p value or, for that matter, with a confidence interval for the indirect effect. This is an issue to which we return below.

Earlier we suggested that researchers may often rely exclusively on index tests to claim mediation. The results of our simulations clearly show that relying only on index tests, particularly the two bias adjusted bootstrap methods, risks Type I errors. This is especially true if one never examines nor reports the individual a_{21} and b_{32} estimates. To survey what in fact is current practice, we reviewed the 2015 volumes of three journals in the field that frequently report mediation analyses (*Journal of Personality and Social Psychology [JPSP]*, *Personality and Social Psychology Bulletin [PSPB]*, and *Psychological Science [PS]*). In that year, there were 97 simple mediation analyses reported in *JPSP*, 130 in *PSPB*, and 66 in *PS*.

Of the 97 reports of simple mediation analyses reported in *JPSP* that year, 82 reported and tested the $a_{21}b_{32}$ product (with the vast majority of these using some form of bootstrap procedures, with specifics often not reported). Of these, there were 60 that also reported the individual a_{21} and b_{32} components and their associ-

ated component tests. Of the 130 simple mediation analyses reported in *PSPB*, 117 reported and tested the $a_{21}b_{32}$ product, using various methods (99 of these used some form of bootstrap procedure). Of these, there were 43 instances in which both components a_{21} and b_{32} , and their associated component tests, were also reported. A similar story emerges for the mediational analyses reported in *PS*. Of the 66 reported analyses, 54 report and test the $a_{21}b_{32}$ product, again primarily using bootstrapping procedures. Of these, 24 report the individual a_{21} and b_{32} estimates and their associated inferential statistics. In total, across the three journals in 2015, 293 separate simple mediation analyses were reported, with the vast majority of these testing the $a_{21}b_{32}$ product with some type of bootstrap procedure. Of these 293 analyses, only 137 reported the individual a_{21} and b_{32} coefficients and their associated standard errors. Thus in 53% of all reported analysis, one has no idea of the magnitude and reliability of the individual a_{21} and b_{32} estimates.

Given our simulation results and those reported by others (Biesanz et al., 2010), statistical power is clearly gained by using one of the bias-adjusted bootstrap methods to test the indirect effect. But these methods also risk inflated Type I error rates unless the individual components of the indirect effect are reported and tested. Our survey of mediation reports in the literature suggests that all too often researchers only report the $a_{21}b_{32}$ product and its test without reporting the individual a_{21} and b_{32} coefficients. When one of these estimates is close to zero, there are substantial risks that index tests of the $a_{21}b_{32}$ product are producing spurious evidence for mediation. Our simulations reveal that unless one examines and tests the individual a_{21} and b_{32} components, the possibility of false mediation claims is far from trivial, particularly when using one of the more popular bias-corrected bootstrap procedures.

Before we draw final conclusions, we examine the generality of our simulation results in more complex mediation cases. To explore this, we ran a series of additional simulations in the cases of two other mediational models for which the index approach has recently been advocated, namely within-participant mediation and moderated mediation.

The Within-Participant Mediation Model

Montoya and Hayes (2017) recently examined mediational analysis in the context of a design where each participant is measured on a dependent variable Y and a mediator M in both of two different conditions. In so doing, these authors revisit an earlier approach proposed by Judd, Kenny, and McClelland (2001) utilizing the following set of models:

$$Y_2 - Y_1 = c_{41} + e_4 \quad (4)$$

$$M_2 - M_1 = a_{51} + e_5 \quad (5)$$

$$Y_2 - Y_1 = c'_{61} + b_{62}(M_2 - M_1) + d_{63}[0.5(M_1 - M_2)] - \overline{0.5(M_1 - M_2)} + e_6 \quad (6)$$

In order to test for the presence of within-participant mediation, Judd et al. (2001) suggest examining and testing the a_{51} and b_{62} coefficients separately, parallel to the component approach in between-participants mediational analysis. In line with that approach, a within-participant indirect effect is said to exist when both coefficients are significant. In sharp contrast, Montoya and Hayes (2017, p. 24) recommend making an inference about the indirect effect as a single index ($a_{51}b_{62}$) and disregarding the component approach proposed by Judd et al. (2001). To examine the relative merits of the component approach and the index approach, we ran a series of simulations.

The Simulations

We adopted the same strategy as for the simple mediation model except that we relied on the more complex Equations 4 and 5. Values of M_1 and Y_1 were sampled from a standard normal population. Values of M_2 were generated from M_1 as a function of a_{51} and a standard normal error. Values of Y_2 were generated from Y_1 as a function of b_{62} times the difference between M_2 and M_1 and a standard normal error. For the sake of the present simulations, c'_{61} and d_{63} were always set at zero.

To simulate a diverse range of situations, the population parameters a_{51} and b_{62} were set to be either nonexistent, small, medium, or large, that is, 0, .14, .39, and .59, respectively. In light of the samples generally encountered in within-participant settings, we retained three different sample sizes, namely 25, 50, and 100. For each of the 48 combinations of conditions, we generated 10,000 samples. We tested for the presence of the indirect effect using the joint-significance test, the Monte Carlo sampling method, and three bootstrap resampling methods, namely the percentile bootstrap, the bias-corrected bootstrap, and the accelerated bias-corrected bootstrap.

Type I Errors

As in the simple mediation simulations, 21 of the 48 populations that we examined were characterized by the absence of an indirect effect because at least one of the two coefficients, a_{51} or b_{62} , was

zero (see Table 3). The relevant numbers for each of the five methods reveal that the Type I errors never exceeds 5% for the joint-significance test. The Monte Carlo sampling method produces more than 5% Type I errors in two cases out of 21, followed closely by the percentile bootstrap method with three cases out of 21. These errors always emerge when coefficient a_{51} is .59.

In contrast, the bias-corrected and the accelerated bias-corrected bootstrap methods produce an excessive number of Type I errors, with no less than 11 cases out of 21 exceeding 5%. Clearly, when one of the two coefficients a_{51} or b_{62} is moderately high, these tests prove too liberal, nearly doubling the proportion of errors when coefficient a_{51} reaches .59.

Power

A total of 27 simulations were characterized by the presence of an indirect effect of varying magnitude because both coefficients were nonzero. In general, power was somewhat higher for the two adjusted bootstrap procedures than for the joint-significance test, the Monte Carlo sampling method, and the percentile bootstrap method. Of these latter three, the power of the joint-significance test was indistinguishable from the other two tests at the larger sample sizes.

Type I Method Inconsistencies

We looked at the relative trustworthiness of the joint-significance test and the Monte Carlo method, on the one hand, and that of the joint-significance test and the two main bootstrap methods, on the other. The two methods in the second cluster consistently showed greater power than those in the former. As can be seen in Table 4, the joint-significance test does slightly better than the Monte Carlo method.

We next compared inconsistencies in Type I errors between pairs of methods. Because the joint-significance test and the percentile bootstrap lead to generally appropriate Type I errors, few cases emerge in which these two tests disagree. Still, when this happens, the simulations indicate that the joint-significance is generally two times more trustworthy, with an average of 38 cases for the joint-significance test and 77 cases for the percentile bootstrap method across the different situations that were examined. The pattern is very different when we compare the joint-significance test and the bias-corrected bootstrap method. Not surprisingly, because the latter method is more liberal, many more situations arise in which this test uniquely leads to a Type I error compared to the joint-significance test. Across the simulated situations, the odds that the joint-significance test proves more trustworthy are in fact 28 to 1. It is noteworthy that, contrary to what was observed in the case of simple mediation, the number of unique Type I errors tends to be larger when it is the a_{51} coefficient that departs from zero ($M = 98$) than when b_{62} does ($M = 67$). This difference is related to the total number of Type I errors.

The Moderated Mediation Model

Muller, Judd, and Yzerbyt (2005) provided the first formal treatment of moderated mediation models (see also Edwards & Lambert, 2007; Fairchild & MacKinnon, 2009; Preacher et al.,

Table 3
Type I Errors (Top Panel) and Power (Bottom Panel) for the Within-Participant Mediation Analysis as a Function of the Population Values of Coefficients a_{51} and b_{62} , Sample Size, and Method

	Type 1 error									
	a = 0		a = .14		a = .39		a = .59		a = .59	
	b = 0	b = .14	b = 0	b = .39	b = 0	b = .59	b = 0	b = .59	b = 0	b = 0
<i>n</i> = 25										
JS	.0020	.0043	.0069	.0144	.0247	.0293	.0391			
MC	.0022	.0044	.0075	.0150	.0277	.0339	.0451			
PB	.0023	.0058	.0089	.0162	.0332	.0347	.0549			
BC	.0075	.0069	.0206	.0394	.0611	.0683	.0880			
ABC	.0096	.0081	.0219	.0421	.0634	.0704	.0897			
<i>n</i> = 50										
JS	.0014	.0071	.0086	.0295	.0373	.0459	.0499			
MC	.0011	.0064	.0072	.0273	.0399	.0503	.0563			
PB	.0020	.0070	.0088	.0293	.0454	.0510	.0638			
BC	.0067	.0176	.0215	.0587	.0806	.0820	.0906			
ABC	.0072	.0187	.0223	.0630	.0836	.0833	.0911			
<i>n</i> = 100										
JS	.0036	.0096	.0153	.0463	.0489	.0469	.0503			
MC	.0028	.0071	.0130	.0450	.0499	.0523	.0551			
PB	.0029	.0080	.0139	.0476	.0559	.0532	.0594			
BC	.0084	.0108	.0338	.0783	.0860	.0707	.0744			
ABC	.0089	.0109	.0342	.0801	.0848	.0691	.0776			
Power										
	a = .14		a = .39		a = .39		a = .59		a = .59	
	b = .14	b = .39	b = .59	b = .14	b = .39	b = .59	b = .14	b = .39	b = .59	b = .59
<i>n</i> = 25										
JS	.0081	.0330	.0642	.0413	.1553	.2949	.0719	.2830	.5240	
MC	.0079	.0336	.0710	.0443	.1673	.3163	.0810	.3076	.5500	
PB	.0103	.0363	.0713	.0502	.1685	.3086	.0899	.3061	.5394	
BC	.0231	.0694	.1223	.0885	.2667	.4222	.1366	.4062	.6435	
ABC	.0255	.0757	.1236	.0918	.2591	.4098	.1374	.3931	.6216	
<i>n</i> = 50										
JS	.0222	.0989	.1552	.1078	.4958	.7084	.1273	.6373	.9149	
MC	.0182	.0972	.1638	.1098	.4944	.7194	.1376	.6542	.9212	
PB	.0186	.0984	.1669	.1134	.4924	.7197	.1463	.6535	.9171	
BC	.0445	.1654	.2199	.1690	.6100	.7924	.1880	.7144	.9407	
ABC	.0468	.1644	.2170	.1693	.5945	.7719	.1847	.7038	.9303	
<i>n</i> = 100										
JS	.0636	.2609	.2840	.2209	.8892	.9701	.2314	.9176	.9986	
MC	.0526	.2613	.2909	.2255	.8926	.9723	.2407	.9226	.9985	
PB	.0554	.2616	.2996	.2289	.8885	.9720	.2528	.9206	.9982	
BC	.1044	.3412	.3414	.2916	.9256	.9793	.2814	.9326	.9987	
ABC	.1029	.3355	.3410	.2910	.9190	.9776	.2813	.9267	.9987	

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap; ABC = accelerated bias-corrected bootstrap. Bold numbers are Type I error rates significantly above .05, i.e., for which .05 falls below the lower limit of the 95% confidence interval.

2007). Consistent with the component approach, Muller et al. (2005) argue that moderated mediation is demonstrated when a moderator, *Z*, significantly moderates at least one path in the causal process linking *X* to *Y* via *M* and when the remaining unmoderated path is also significantly different from zero (see Muller et al., 2005, for a full presentation). In contrast, Hayes (2013, 2015) argues that a single test of a product of regression coefficients, the so-called index of moderated mediation, should serve as a formal test of moderated mediation. As Hayes

(2015) notes, relying on this test allows one to disregard the fact that the examined data fail to reveal the presence of a significant interaction between any variable in the model and the moderator. In other words, an indirect effect could be moderated even if one cannot show significant moderation of either of its components.

As earlier, we wanted to compare the performance of the component approach and the index approach. Because a variety of patterns correspond to a situation of moderated mediation, we

Table 4

Type I Errors Uniquely Due to One as Opposed to the Other Method as a Function of the Values of Coefficients and, Pair of Methods, and Sample Size in the Context of a Within-Participant Mediation Analysis

Comparison	n	a = 0	a = 0	a = .14	a = 0	a = .39	a = 0	a = .59	Sub-Mean	Mean
		b = 0	b = .14	b = 0	b = .39	b = 0	b = .59	b = 0		
JS = 1 & MC = 0	25	2	7	10	21	18	28	36	17.43	23.33
	50	4	16	29	47	32	31	33	27.43	
	100	11	28	26	54	22	14	21	25.14	
JS = 0 & MC = 1	25	4	8	16	27	4	74	13	20.86	36.76
	50	1	9	59	25	58	75	43	38.57	
	100	3	3	86	41	86	68	69	50.86	
JS = 1 & PB = 0	25	6	11	16	34	57	55	85	37.71	37.57
	50	4	14	22	55	64	35	79	39.00	
	100	10	23	40	41	66	16	56	36.00	
JS = 0 & PB = 1	25	9	26	36	52	142	109	243	88.14	77.05
	50	10	13	24	53	145	86	218	78.43	
	100	3	7	26	54	136	79	147	64.57	
JS = 1 & BC = 0	25	0	3	1	6	19	11	29	9.86	8.86
	50	1	2	4	4	11	5	20	6.71	
	100	0	3	4	2	15	5	41	10.00	
JS = 0 & BC = 1	25	55	96	138	256	383	401	518	263.86	251.67
	50	54	107	133	296	444	366	427	261.00	
	100	58	131	189	322	386	243	282	230.14	

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap. A 1 means that the method led to the decision of the presence of an indirect effect and a 0 means that the method led to the decision of an absence of indirect effect. For instance, when a = 0 and b = 0 and JS = 1 and MC = 0 and n = 50, 2 means that on two occasions (out of 10,000) the joint-significance test led to conclude in favor of an indirect effect when the Monte Carlo sampling method did not.

decided to focus on what is known as first-stage moderated mediation (Edwards & Lambert, 2007). In this situation (see Equations 7 and 8), the effect of X on M is moderated by Z, but the partial effect of M on Y is unmoderated.

$$M = b_{70} + a_{71}X + a_{72}Z + a_{73}XZ + e_7 \tag{7}$$

$$Y = b_{80} + c'_{81}X + b_{82}M + e_8 \tag{8}$$

As Edwards and Lambert (2007) and Preacher, Rucker, and Hayes (2007) show, the indirect effect in this situation corresponds to the product of the conditional effect of X on M from Equation 7 times the effect of M on Y from Equation 8, $(a_{71} + a_{73}Z)b_{82}$, or equivalently $a_{71}b_{82} + a_{73}b_{82}Z$. Because $a_{73}b_{82}$ quantifies the impact of Z on the indirect of X on Y, Hayes (2015) calls it the index of moderated mediation, at least for the first stage (and direct) moderated mediation model. Comparable products can be computed for other models (see Hayes, 2015, for a detailed discussion). It should be clear by now that the component approach would recommend evaluating the statistical significance of both coefficients making up the index, namely a_{73} and b_{82} . A moderated, or conditional, indirect effect would be supported only when both coefficients are found to be different from zero. We conducted simulations to try and clarify the relative performance of various methods relying on one or the other approach.

The Simulations

The simulations we conducted compared the performance of the same five methods as above in the context of a first stage moderated mediation model. To this end, we relied on Equations 7 and 8. Values of X and Z were sampled from a standard normal population and centered around the sample means before computing their product. Values of M were generated from XZ multiplied

by the population path a_{73} and adding a standard normal error. Values of Y were generated using the population path b_{82} times M and adding a standard normal error. In other words, for the sake of the present simulations, c'_{81} as well as a_{71} and a_{72} were always set at zero, meaning that, in the population, only an interaction effect influenced M and no direct effect of X affected Y. As before, the standard normal errors were added to produce sampling discrepancy between the population parameters and their estimates.⁵

To simulate a diverse range of situations, the population parameters a_{73} and b_{82} were set to be either nonexistent, small, medium, or large, that is, 0, .14, .39, and .59, respectively. Given the samples usually studied in a majority of psychology fields, we opted for three different sample sizes, namely 50, 100, and 200. As before, for each of the 48 combinations of conditions, we generated 10,000 samples. We tested for the presence of the conditional indirect effect using the joint-significance test, the Monte Carlo sampling method, and three bootstrap resampling methods, namely the percentile bootstrap, the bias-corrected bootstrap, and the accelerated bias-corrected bootstrap.

Type I Errors

Out of the 48 situations examined, 21 were defined by an absence of a so-called conditional indirect effect because at least one of the

⁵ As can be seen in Equation 8, the Stage 2 moderation is zero in the present simulations. This corresponds to the situation examined by Hayes (2017; see also Preacher et al., 2007). It should be noted that in any given dataset, one does not know whether the other stage moderation is truly zero. Therefore, contrary to what is done in the PROCESS macro, we recommend estimating both stage moderations, in line with Muller et al. (2005). As detailed in the online supplemental material, the JSmediation package is consistent with this full model approach.

two coefficients, a_{73} and b_{82} , was zero. Table 5 shows the numbers of Type I errors for each of the five methods. It can be seen that, regardless of sample size and parameter values, the prevalence of Type I errors for the joint-significance test significantly exceeds .05 only once. This pattern is not true for the four index methods. When one of the parameter values was either moderate or large and the other one was zero, the bias-corrected and the adjusted bias-corrected bootstrap procedures yielded empirical Type I error rates that were almost always larger than .05. The other two index methods (Monte Carlo and percentile bootstrap) yielded more appropriate Type I error rates, but even these were sometimes above the .05 value. In sum,

only the joint-significance test showed essentially no inflation in Type I error rates.

Power

In total, 27 situations were characterized by the presence of an indirect effect of varying magnitude because both coefficients were nonzero (see Table 5). Again, two clusters of methods can be distinguished, with the joint-significance test, the Monte Carlo sampling method, and the percentile bootstrap method, on the one hand, and the two remaining methods, on the other. The propor-

Table 5
Type I Errors (Top Panel) and Power (Bottom Panel) for the Moderated Mediation Analysis as a Function of the Population Values of Coefficients a_{73} and b_{82} , Sample Size, and Method

		Type 1 error									
		a = 0 b = 0	a = 0 b = .14	a = .14 b = 0	a = 0 b = .39	a = .39 b = 0	a = 0 b = .59	a = .59 b = 0			
<i>n</i> = 50		JS	.0035	.0080	.0069	.0382	.0352	.0475	.0466		
		MC	.0029	.0066	.0053	.0395	.0334	.0534	.0503		
		PB	.0027	.0070	.0055	.0386	.0323	.0510	.0510		
		BC	.0068	.0158	.0119	.0618	.0544	.0703	.0734		
		ABC	.0075	.0188	.0132	.0688	.0485	.0767	.0623		
<i>n</i> = 100		JS	.0018	.0140	.0134	.0451	.0479	.0497	.0509		
		MC	.0011	.0106	.0107	.0467	.0493	.0537	.0537		
		PB	.0017	.0125	.0099	.0537	.0516	.0560	.0607		
		BC	.0054	.0294	.0253	.0783	.0788	.0683	.0734		
		ABC	.0059	.0300	.0259	.0810	.0641	.0737	.0612		
<i>n</i> = 200		JS	.0022	.0230	.0257	.0537	.0548	.0494	.0518		
		MC	.0012	.0189	.0205	.0549	.0581	.0513	.0556		
		PB	.0019	.0219	.0209	.0625	.0590	.0575	.0574		
		BC	.0056	.0454	.0453	.0777	.0746	.0628	.0637		
		ABC	.0064	.0472	.0425	.0782	.0596	.0679	.0526		
		Power									
		a = .14 b = .14	a = .14 b = .39	a = .14 b = .59	a = .39 b = .14	a = .39 b = .39	a = .39 b = .59	a = .59 b = .14	a = .59 b = .39	a = .59 b = .59	
<i>n</i> = 50		JS	.0245	.1096	.1463	.1192	.5382	.6224	.1851	.7915	.9150
		MC	.0228	.1104	.1587	.1208	.5417	.6751	.1937	.7982	.9229
		PB	.0196	.0979	.1374	.1070	.4797	.6110	.1846	.7538	.8851
		BC	.0366	.1447	.1742	.1549	.5719	.6804	.2339	.8182	.9155
		ABC	.0385	.1489	.1788	.1393	.5300	.6421	.1965	.7568	.8860
<i>n</i> = 100		JS	.0712	.2590	.2607	.2946	.9258	.9394	.3480	.9884	.9981
		MC	.0627	.2611	.2719	.2970	.9282	.9429	.3596	.9890	.9983
		PB	.0644	.2535	.2664	.2913	.9127	.9282	.3642	.9863	.9971
		BC	.1125	.3180	.2978	.3627	.9409	.9435	.3951	.9902	.9979
		ABC	.1085	.3133	.2938	.3227	.9180	.9280	.3478	.9829	.9958
<i>n</i> = 200		JS	.2461	.4849	.4906	.5618	.9994	.9999	.6154	1.0000	1.0000
		MC	.2239	.4940	.5002	.5688	.9994	.9991	.6229	1.0000	1.0000
		PB	.2214	.4876	.4933	.5687	.9989	.9984	.6234	.9999	1.0000
		BC	.3245	.5132	.5133	.6119	.9993	.9987	.6400	.9999	1.0000
		ABC	.3042	.5163	.5038	.5630	.9986	.9980	.5876	.9999	1.0000

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap; ABC = accelerated bias-corrected bootstrap. Bold numbers are Type I error rates significantly above .05, i.e., for which .05 falls below the lower limit of the 95% confidence interval.

tions of positive decisions suggest greater power for the methods in the latter cluster. In contrast, the three other methods are more conservative. The percentile bootstrap method proves slightly more powerful than the joint-significance test and the Monte Carlo sampling method.

Type I Method Inconsistencies

We compared the trustworthiness of the joint-significance method to that of the two main bootstrap methods. As can be seen in Table 6, the joint-significance test performs slightly better than the Monte Carlo method. Also, because the joint-significance test and the percentile bootstrap lead to a limited number of erroneous decisions that an indirect effect is present when in fact there is none, a small number of cases emerge whereby one of these two tests errs when the other does not. However, when this happens, the joint-significance test is globally more trustworthy, with an average of 74 cases for the joint-significance test and 95 cases for the percentile bootstrap method across the different situations that were examined in the simulations. The pattern is very different when we compare the joint-significance test and the bias-corrected bootstrap method. Because the latter method is more liberal, many more situations arise in which this test uniquely leads to a Type I error compared with the joint-significance test. The odds that the joint-significance test proves more trustworthy are 6 to 1 across the simulated situations.

To sum up, the simulations show that the pattern of the Type I errors, power, and inconsistencies is the same as the one observed for the simulations using the simple mediation model. It thus appears that here too claiming moderated mediation is safer, in terms of avoiding Type I errors, when relying on the component approach rather than the index approach.

Recommendations

Both recent recommendations and practice for those claiming mediation have been to rely on a single mediation index and test whether its bootstrap-based confidence interval excludes zero (Hayes, 2013). The wide availability of stand-alone or software-embedded macros has made this strategy easy and appealing. The simulations we conducted for simple mediation, within-participant mediation, and moderated mediation demonstrate the misconceptions and dangers of relying solely on such an approach.

One misconception is that a component approach, which requires that both of two coefficients be significant to demonstrate mediation, leads to an inflation of Type I errors. This is simply not the case since the requirement is that they each be significant. In fact, our simulations confirm that the joint-significance test is the only one holding alpha at appropriate levels with only one instance in which the empirical Type I error rate exceeded 5% (across 63 studied situations). The most frequently used index method (i.e., the bias-corrected bootstrap, which was until recently the default approach in available macros) has substantially inflated alpha levels, particularly in cases where one of the two components of the indirect effect is in fact zero and the other is relatively large. As for power, our simulations reveal that the joint-significance test performs as well as other methods, except for the more liberal bias-corrected and accelerated bias-corrected bootstrap methods. In light of these findings, the joint-significance method constitutes the best compromise between Type I error rate and power and ought to be the method of choice (see Table 7).

Reliance on index tests unfortunately means that researchers may not even look at, let alone test, the components of the indirect effect. Indeed, our review of recently published reports of mediation shows that researchers very often fail to even report the

Table 6
Type I Errors Uniquely Due to One as Opposed to the Other Method as a Function of the Values of Coefficients a_{73} and b_{82} , Pair of Methods, and Sample Size in the Context of a Moderated Mediation Analysis

Comparison	n	a = 0	a = 0	a = .14	a = 0	a = .39	a = 0	a = .59	Sub-Mean	Mean
		b = 0	b = .14	b = 0	b = .39	b = 0	b = .59			
JS = 1 & MC = 0	50	6	18	21	44	51	18	26	26.29	30.67
	100	7	44	34	36	39	20	24	29.14	
	200	10	55	61	35	31	31	33	36.57	
JS = 0 & MC = 1	50	12	4	5	57	33	77	63	35.86	35.29
	100	0	10	7	52	53	60	53	33.57	
	200	0	14	9	47	64	50	71	36.43	
JS = 1 & PB = 0	50	20	38	31	145	109	155	129	89.57	73.71
	100	9	57	59	94	92	120	50	68.71	
	200	9	70	80	82	61	88	50	62.86	
JS = 0 & PB = 1	50	12	28	17	149	80	190	169	92.14	95.43
	100	8	42	24	180	129	183	148	102.00	
	200	6	59	32	170	103	169	106	92.14	
JS = 1 & BC = 0	50	8	20	14	78	58	81	64	46.14	33.62
	100	3	15	14	31	22	78	28	27.29	
	200	2	14	13	36	23	67	37	27.43	
JS = 0 & BC = 1	50	41	98	64	314	250	309	332	201.14	204.67
	100	39	169	133	363	331	265	253	221.86	
	200	36	238	209	276	221	201	156	191.00	

Note. JS = joint-significance test; MC = Monte Carlo sampling method; PB = percentile bootstrap; BC = bias-corrected bootstrap. A 1 means that the method led to the decision of the presence of an indirect effect and a 0 means that the method led to the decision of an absence of indirect effect. For instance, when a = 0 and b = 0 and JS = 1 and MC = 0 and n = 50, 6 means that on six occasions (out of 10,000) the joint-significance test led to conclude in favor of an indirect effect when the Monte Carlo sampling method did not.

Table 7
Type I Error and Power Associated With the Various Methods

	Type I error	Power
Joint-significance test ¹	Very good performance	Good performance
Monte Carlo	Good performance	Good performance
Percentile bootstrap	Good performance	Good performance
Bias-corrected bootstrap	Bad performance ²	Very good performance
Accelerated bias-corrected bootstrap	Bad performance ²	Very good performance

¹ The joint-significance test offers the best balance between Type I error rate and power. ² The Type I error rate is substantially inflated when on component path of the indirect path is in fact zero and the other component path is large.

magnitude of the indirect effect components when relying on index tests of its significance. The failure to critically examine these components, in our opinion, has possibly led to unwarranted claims of mediation that may not be replicable. This is particularly likely to be the case when one of the two component effects is especially large. Then that effect may by itself lead to a relatively large and significant mediation index. In simple mediation, this is especially likely to happen either when the mediator is but a manipulation check (leading to a large a_{21} effect) or when it is essentially an alternative measure of the dependent variable (leading to a large b_{32} effect). In both situations, claims of causal mediation are dubious.

In a thoughtful contribution, Fiedler, Schott, and Meiser (2011) similarly stress the difficulties inherent to these two situations. These authors take the example of a researcher who wants to conduct a study inspired by the elaboration likelihood model showing that the quality of arguments (the independent variable X) affects attitude change (the dependent variable Y) through recipients' cognitive responses (the mediator M). Although the researcher's mediational model may come as entirely warranted on theoretical grounds and can in fact be borne out in the collected data, a closer consideration of the individual components may shed interesting light on the viability of the hypothesized causal model. Indeed, one could argue that the mediator, M , is simply a second measure of the resulting attitude, Y . In other words, the supporting thoughts and counterthoughts simply represent an alternative measure of the attitude change induced by the attitude strength manipulation. Because in this case both M and Y are interchangeable, and indeed highly correlated, consequences of X , the b path would be highly significant. Only looking at the significant indirect effect would likely be questionable in this situation. Alternatively, it could also be that M is simply another reflection of the independent variable X . After all, one way to come up with a strong versus weak message is to pretest the thoughts triggered by a series of arguments and then create the two messages accordingly. From this perspective, the cognitive responses are but a reflection of the same construct as the one underlying X . This situation would make for a high a component and again lead to a highly significant indirect effect, creating the same difficulty with respect to the causal claims of the researcher.

In light of the above, our recommendations are straightforward (see Table 8). Claims of mediation should properly be guided by the component approach and be based on joint-significance tests to avoid spurious mediation claims. At the same time, given the presence of mediation, one should then follow up with appropriate examinations of the magnitude of the effect, using resampling

methods to examine the confidence interval of the overall indirect effect. With the exception of the accelerated bias-corrected (see Biesanz et al., 2010) and bias-corrected (the present simulations) methods that are decidedly too liberal, any resampling method would seemingly do the job. In light of our simulations for simple mediation, within-participant mediation, and moderated mediation, our preference is for the Monte Carlo method, because it is least likely to yield inconsistencies with the joint-significance test.

An Illustrative Example and a Dedicated Package

Let us illustrate the recommended analytic strategy using the simple mediation example presented in the introduction (Ho et al., 2017). Remember that participants were found to report more hypodescent when informed that Black-White biracials do versus do not experience discrimination, $c_{11} = 0.17$, $p = .04$. In line with the present recommendations, one would first want to test the significance of the a_{21} path linking the independent variable to the mediator and the significance of the b_{32} path linking the mediator to the dependent variable. As predicted by the author, the high discrimination condition increased the perception of linked fate, $a_{21} = 0.77$, $p < .0001$, and the perception of linked fate led to the more hypodescent, $b_{32} = 0.19$, $p < .001$. Having established the presence of an indirect effect by means of the significance of both individual components (i.e., the joint-significance test), one would then proceed with Monte Carlo resampling to compute the confidence interval for the indirect effect, the product of these two estimated components. To do so, one would rely on the value of these two coefficients along with their respective standard errors, $sea = 0.085$ and $seb = 0.033$. Using 10,000 samples, the mean value of the indirect effect equals 0.14, which corresponds to the difference between c_{11} and c'_{31} , 0.17 and 0.03, respectively, with a 95% confidence interval ranging from 0.09 to 0.21.

These analyses can be conducted very easily with the JSmediation R package. After creating a contrast code for the condition

Table 8
Recommendations for the Analysis of Mediation

Step 1	Examination of the component paths by means of joint-significance test if all component paths of the indirect effect are significant, then conclude in favor of mediation and proceed.
Step 2	Examination of the magnitude and confidence interval of indirect effect by means of any resampling method (preferably Monte Carlo resampling method).

variable (using -0.5 and $+0.5$ for the low discrimination and high discrimination conditions, respectively, thanks to the `build_contrast` function), one would use the `mdt_simple` function, that is, the function for simple mediation. Using this function provides a direct test of the c_{11} , c'_{31} , a_{21} , and b_{32} paths. Next, using the `add_index` function gives access to a point estimate for the ab indirect effect as well as the Monte Carlo 95% confidence.

Issues of Power

In this article, we compared the power performance of various methods. Our simulations revealed that the joint-significance test proved quite satisfactory compared with other methods testing for the presence of an indirect effect whether in a situation of simple mediation, of within-participant mediation or moderated mediation. In general, researchers should indeed consider the level of power they would like to secure before they collect the data. Interested readers can definitely build upon several recent contributions dealing with this topic to help make their decision in this respect (Fritz & MacKinnon, 2007; Kenny & Judd, 2014; Loeys, Moerkerke, & Vansteelandt, 2015; Preacher & Kelley, 2011).

Assumptions of Normality

One potential limitation of the joint-significance test we are recommending is that it relies on normal distribution theory assumptions, whereas the tests of the $a_{21}b_{32}$ index rely instead on nonparametric methods that may perform more adequately in the presence of outliers and other violations of the assumption that residuals in the mediation models have normal distributions. The original justification for using nonparametric methods for testing the $a_{21}b_{32}$ index is that it is known that the product of coefficients does not have a normal sampling distribution. This is true even when normal distribution assumptions can be met for the models' residual errors. Thus, index approaches were not originally justified based on their ability to deal more appropriately with "nasty and unruly data" (McClelland, 2014).

We would like to stress that the presence of data (or more precisely models' residuals) that violate normal distribution theory assumptions is an issue that is orthogonal to whether a component or an index approach is used to test for mediation. At the same time, a common intuition here is that nonparametric approaches such as bootstrapping are likely to be more appropriate than using standard errors that depend on normal distribution assumptions in order to test null hypotheses on the individual components. Accordingly, we would recommend that bootstrapping approaches be used in examining and evaluating the mediation components individually when it is known that model residuals are likely to violate normal distribution assumptions. Even in such cases, we would recommend that confidence intervals for the individual component coefficients be examined and reported rather than reporting only the ab indirect effect and its associated bootstrap-derived confidence interval.

As we mentioned earlier, Biesanz et al. (2010) also investigated the robustness of various mediation methods, among which were the joint-significance test, the percentile bootstrap method, and the accelerated bias-corrected bootstrap method, in the presence of non-normal residuals. Interestingly, their simulations reveal that the accelerated bias-corrected bootstrap method (and, we would

venture in light of our own simulations, the bias-corrected bootstrap method as well) displays excessive Type I error rates, even with samples as large as $N = 500$. Both the joint-significance test and the percentile bootstrap perform satisfactorily, with a small advantage to the former. In short, based on Biesanz et al.'s (2010) efforts, it appears that, even without any correction for non-normality, the component approach constitutes a sound analytic strategy to evaluate the presence of an indirect effect.

Concluding Thoughts

In the end, claims for mediation, whether in the simple case, in the within-participant case, or in the moderated case, all depend on the plausibility of the entirety of the mediational model. Such claims necessarily involve a close inspection of the estimated coefficients of the model. Examining and testing only a single index of mediation, the indirect effect, risks not only committing Type I errors, but also failing to understand what the underlying model really signifies. Only after the individual components of the indirect effect are shown to support researchers' claims, should one use resampling methods to compute a confidence interval for the indirect effect. The tests of the individual components are used to argue for the significance of the indirect effect. The confidence interval reveals its magnitude.

References

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182. <http://dx.doi.org/10.1037/0022-3514.51.6.1173>
- Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . . Johnson, V. (2017). *Redefine statistical significance*. Retrieved from osf.io/preprints/psyarxiv/mky9j
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research, 45*, 661–701. <http://dx.doi.org/10.1080/00273171.2010.498292>
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29. <http://dx.doi.org/10.1177/0956797613504966>
- Dohle, S., & Siegrist, M. (2014). Fluency of pharmaceutical drug names predicts perceived hazardousness, assumed side effects and willingness to buy. *Journal of Health Psychology, 19*, 1241–1249. <http://dx.doi.org/10.1177/1359105313488974>
- Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*, 1–22. <http://dx.doi.org/10.1037/1082-989X.12.1.1>
- Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*, 87–99. <http://dx.doi.org/10.1007/s1121-008-0109-6>
- Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47*, 1231–1236. <http://dx.doi.org/10.1016/j.jesp.2011.05.007>
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239. <http://dx.doi.org/10.1111/j.1467-9280.2007.01882.x>
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47*, 61–87. <http://dx.doi.org/10.1080/00273171.2012.640596>

- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*, 1–22. <http://dx.doi.org/10.1080/00273171.2014.962683>
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York, NY: Guilford Press.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927. <http://dx.doi.org/10.1177/0956797613480187>
- Ho, A. K., Kteily, N. S., & Chen, J. M. (2017). “You’re one of us”: Black Americans’ use of hypodescent and its association with egalitarianism. *Journal of Personality and Social Psychology*, *113*, 753–768. <http://dx.doi.org/10.1037/pspi0000107>
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619. <http://dx.doi.org/10.1177/0193841X8100500502>
- Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*, 115–134. <http://dx.doi.org/10.1037/1082-989X.6.2.115>
- Judd, C. M., Yzerbyt, V. Y., & Muller, D. (2014). Mediation and moderation. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 653–676). Cambridge, UK: Cambridge University Press.
- Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science*, *25*, 334–339. <http://dx.doi.org/10.1177/0956797613502676>
- Krueger, J. I., & Heck, P. R. (2017). The heuristic value of p in inductive statistical inference. *Frontiers in Psychology*, *8*, 908. <http://dx.doi.org/10.3389/fpsyg.2017.00908>
- Loeys, T., Moerkerke, B., & Vansteelandt, S. (2015). A cautionary note on the power of the test for the indirect effect in mediation analysis. *Frontiers in Psychology*, *5*, 1549. <http://dx.doi.org/10.3389/fpsyg.2014.01549>
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Routledge.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*, 384–389. <http://dx.doi.org/10.3758/BF03193007>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104. <http://dx.doi.org/10.1037/1082-989X.7.1.83>
- Mackinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128. http://dx.doi.org/10.1207/s15327906mbr3901_4
- Mackinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41–62. http://dx.doi.org/10.1207/s15327906mbr3001_3
- McClelland, G. H. (2014). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 608–626). Cambridge, UK: Cambridge University Press.
- Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*, 6–27. <http://dx.doi.org/10.1037/met0000086>
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863. <http://dx.doi.org/10.1037/0022-3514.89.6.852>
- Ong, D. C. (2014). *A primer to bootstrapping and an overview of doBootstrap*. Retrieved from <https://web.stanford.edu/class/psych252/tutorials/doBootstrapPrimer.pdf>
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93–115. <http://dx.doi.org/10.1037/a0022658>
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227. <http://dx.doi.org/10.1080/00273170701341316>
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77–98. <http://dx.doi.org/10.1080/19312458.2012.679848>
- R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology, 1982* (pp. 290–312). Washington, DC: American Sociological Association. <http://dx.doi.org/10.2307/270723>
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*, 692–700. <http://dx.doi.org/10.3758/s13428-011-0076-x>

Received January 16, 2018
 Revision received July 11, 2018
 Accepted July 15, 2018 ■

Correction to Brienza et al. (2017)

In the article “Wisdom, Bias, and Balance: Toward a Process-Sensitive Measurement of Wisdom-Related Cognition” by Justin P. Brienza, Franki Y. H. Kung, Henri C. Santos, D. Ramona Bobocel, and Igor Grossmann (*Journal of Personality and Social Psychology*, Advance online publication, September 21, 2017. <http://dx.doi.org/10.1037/pspp0000171>), the original supplemental has been revised to include a clarifying note to the Tests of model fit over larger sample (Samples C–G) section and post-peer review analyses added to the Post-peer review Factor Analytic Tests section.

All versions of this article have been corrected.

<http://dx.doi.org/10.1037/pspp0000234>