

Meaningful Mediation Analysis:  
Plausible Causal Inference and Informative Communication

RIK PIETERS

Forthcoming, Journal of Consumer Research

Rik Pieters ([pieters@tilburguniversity.nl](mailto:pieters@tilburguniversity.nl)) is Arie Kapteijn Professor of Economics and Management, Department of Marketing, Tilburg School of Economics and Management, Tilburg University, Box 90153, 5000 LE Tilburg, the Netherlands. I thank Sara Kim, Ann McGill, Danielle Kapor, and Zakary Tormala for sharing the data of their research. I am also grateful to Derek Rucker and Stephen Spiller for their outstanding—and at the time anonymous—associate editor and reviewer reports, and the editorial team for their support and suggestions. Thanks are due as well to seminar participants at Leuven University and the Free University of Amsterdam, and in particular to Esther Jaspers, Ana Martinovici, Niels van de Ven, Stijn van Osselaer, and Joachim Vosgerau for helpful comments on earlier versions. Martin Schreier proposed the Sweetspot name. Hannes Datta wrote the code to summarize results from MplusAutomation. Supplementary material on <https://github.com/RikPieters/Meaningful-Mediation>.

This tutorial was invited by editors Darren Dahl, Eileen Fischer, Gita Johar, and Vicki Morwitz. Derek Rucker served as associate editor on this tutorial. The tutorial underwent a modified review process wherein it was sent to one reviewer and one associate editor for a single round.

## ABSTRACT

Statistical mediation analysis has become the technique of choice in consumer research to make causal inferences about the influence that a treatment has on an outcome via one or more mediators. This tutorial aims to strengthen two weak links that impede statistical mediation analysis from reaching its full potential. The first weak link is the path from mediator to outcome, which is correlation. Six conditions are described that this correlation needs to meet in order to make plausible causal inferences: directionality, reliability, unconfoundedness, distinctiveness, power, and mediation. Recommendations are made to increase the plausibility of causal inferences based on statistical mediation analysis. Sweetspot analysis is proposed to establish whether an observed mediator-outcome correlation falls within the region of statistically meaningful correlations. The second weak link is the communication of mediation results. Four components of informative communication of mediation analysis are described: effect decomposition, effect size, difference testing, and data sharing. Recommendations are made to improve the communication of mediation analysis. A review of 166 recently published mediation analyses in the *Journal of Consumer Research*, a re-analysis of two published datasets, and Monte Carlo simulations support the conclusions and recommendations.

*Keywords:* mediation analysis, causal inference, experiment, bias, knowledge accumulation

## INTRODUCTION

Mediation analysis is done to make causal inferences about the influence that a treatment has on an outcome via one or more mediators. It holds the promise of contributing to improved theories about the causal processes that account for treatment effects, and thereby to more effective and efficient treatments. Seminal publications on its foundations, and the availability of tailored statistical procedures have made mediation analysis an indispensable technique in the researcher's toolbox (Baron and Kenny 1986; Hayes 2012; MacKinnon 2008; Preacher, Rucker, and Hayes 2007; Shrout and Bolger 2002; Zhao, Lynch, and Chen 2010). Mediation analysis is considered: "...almost mandatory for new social-psychology manuscripts" (Bullock, Green, and Ha 2010, p. 550). This holds for consumer research as well. Of the 121 articles based on experiments in volumes 41 and 42 (2014-2016) of the *Journal of Consumer Research*, 71% contained at least one mediation analysis.

Mediation analysis is not a prerequisite for each and every empirical research report, on the contrary. Establishing replicable effects of treatments on outcomes may be important by itself, and there are alternative, sometimes superior approaches to providing process evidence. Yet, if research relies on mediation analysis to make inferences about causal processes, its findings need to be meaningful and meaningfully reported. This tutorial aims to contribute to such meaningful mediation analysis, by strengthening two weak links in the foundation and application of mediation analysis. The first weak link concerns the plausibility of causal inferences from mediation analysis. Mediation analysis rests on the path between the mediator and outcome, which is a correlation. The path needs to meet six conditions to permit plausible causal inferences. Failure to meet the conditions can lead to large and largely intractable biases that can render causal inferences from mediation analysis meaningless. As yet, applications of

mediation analysis in the consumer behavior literature mostly consider the final condition only, as shown later. The second weak link concerns the contribution of mediation analysis results to theory and practice. Insight into the causal processes of interest, and knowledge accumulation across mediation analyses critically rests on comprehensive communication of the results of mediation analysis. As yet, such comprehensive communication is rare in the consumer behavior literature, as shown later. This tutorial aims to strengthen these two weak links in mediation analysis. Throughout, it provides examples of best practices in the consumer behavior literature.

The first section describes three common mediation models: basic, multiple, and moderated mediation. The second section describes six conditions that the path between mediator and outcome needs to meet. It identifies potential biases when conditions are not met and offers recommendations for improvement. The second section also introduces Sweetspot analysis, which explores the region of statistically meaningful correlations between mediator and outcome. The third section describes four components of comprehensive communication of mediation analysis, and offers recommendations for improvement. The final section draws conclusions. The tutorial builds on a review of 166 mediation analyses published in the *Journal of Consumer Research* (2014-2016), a re-analysis of two published mediation analyses (Kim and McGill 2011; Kupor and Tormala 2015), and Monte Carlo simulations. Supplementary material on the review, re-analyses, and simulations is available on-line.

## MEDIATION ANALYSIS

Mediation analysis decomposes the total effect that an input variable (X) has on an outcome variable (Y) into an indirect effect that is transferred via a mediator (M) and a conditional direct effect. The focus here is on natural or controlled experiments with random

assignment of participants to one or more treatment and control conditions. The terms treatment (X), mediator (M), and outcome (Y) denote the three key variables. The diamond in figure 1 indicates that X is an experimental variable, from now on called “treatment.” Boxes indicate that M and Y are observational variables. Figure 1 summarizes three mediation models: basic mediation, multiple mediation, and moderated mediation.

In volumes 41 and 42 of the *Journal of Consumer Research* (2014-2016), 121 of the 138 articles included experiments (88%), and from these 86 (71%) contained at least one mediation analysis, with an average of 1.93. Of the 166 relevant cases, 55 (33%) examined basic mediation, 29 (17%) multiple mediation, and 82 (49%) moderated mediation. Combined multiple and moderated mediation analyses were coded as moderated mediation in, from now on, the “mediation review.” Table 1 has a summary.

\*\*\* Insert figure 1 and table 1 \*\*\*

In the context of path analysis, Wright (1921) was the first to propose that the indirect effect of one variable on another is captured by the product of the path weights connecting the two variables. Mediation analysis relies on this idea. It decomposes the total treatment effect into an indirect effect, which is the product-of-coefficients of the paths leading from the treatment to the outcome, and the conditional direct effect which is the remaining treatment effect. The three mediation models differ in how the total treatment effect is decomposed.

Model 1 in figure 1 is the basic mediation model. It has a single X, M, and Y. The total treatment effect (c) is decomposed into an indirect effect ( $a \times b$ , or  $ab$ ) and a conditional direct effect ( $c'$ ). Thus,  $c = a \times b + c'$ .

Model 2 in figure 1 is a multiple mediation model, here with two mediators. The total treatment effect is decomposed into a conditional direct effect and several indirect effects. In

multiple mediation models all paths between treatment, mediators, and outcomes are estimated to appropriately decompose the total treatment effect (Preacher and Hayes 2008). Parallel mediation and serial (or sequential) mediation are different versions of multiple mediation. It is rarely appreciated that the difference between various multiple mediation models critically hinges on the hypothesized direction of a single path only, namely between the two mediators (M1 and M2), which is called the d-link here. Figure 1 shows three specifications of the direction of the d-link (d1 to d3). Each d-link (1 to 3) is included in only one of the three multiple mediation models: d1 in Model 2.1, d2 in Model 2.2, and d3 in Model 2.3, as shown next. Although each specification of the d-link produces a theoretically distinct model, the models are statistically equivalent. A later section examines its implications for causal inference.

Model 2.1 is a parallel mediation model. The treatment effect runs via two separate, but correlated, pathways to the outcome, namely from X via M1 to Y, and from X via M2 to Y. This model has the d1-link, which is the covariance between the residuals of the two mediators (M1 and M2). Thus, the d-link is undirected in this model. The *total indirect* treatment effect in this model is then the sum of the two parallel paths ( $a_1 \times b_1 + a_2 \times b_2$ ).

Model 2.2 is a serial mediation model. The treatment effect runs from X via M1 to M2 and then to Y. This model has the d2-link from M1 to M2. The hypothesized serial indirect effect is the product of the respective paths, namely  $a_1 \times d_2 \times b_2$ . The *total indirect* treatment effect is the sum of the serial indirect and the two other indirect effects:  $a_1 \times d_2 \times b_2 + a_1 \times b_1 + a_2 \times b_2$ .

Model 2.3 is an alternative serial mediation model. The treatment effect now runs from X via M2 to M1 and then to Y. This model estimates the d3-link from M2 to M1. The hypothesized serial indirect effect is the product of the respective paths:  $a_2 \times d_3 \times b_1$ . The *total indirect* treatment effect is the sum of the serial indirect and the two other indirect effects:  $a_2 \times d_3 \times b_1 +$

$a_1 \times b_1 + a_2 \times b_2$ . Of the 29 multiple mediation models in the mediation review, 13 (45%) were parallel and 16 (55%) were serial, with blends between the two model types coded as serial.

Model 3 in figure 1 is a moderated mediation model. In moderated mediation, the indirect treatment effect is conditional on a moderating variable (W). The moderator can influence the treatment effect on the mediator (first-stage moderation), the correlation between mediator and outcome (second-stage moderation) or both. Another type of moderation occurs when the treatment influences the mediator, and also moderates the relationship between mediator and outcome (Judd and Kenny 1981; Preacher, Rucker, and Hayes 2007; Valeri and VanderWeele 2013). None of the analyses in the mediation review examined it. The moderator can be experimental (diamond) or observational (box). Model 3 in figure 1 shows first-stage moderation: 96% of the 82 cases in the mediation review with a moderated mediation examined it. The conditional direct effects of the treatment (X) and moderator (W) on the outcome (Y) are accounted for as well. The critical test is then whether the interaction between X and W has an effect via the mediator (a3-path) on the outcome (b-path):  $a_3 \times b$ . There is evidence for moderated mediation when this conditional indirect effect is statistically significant.

Bollen and Stine (1990) first proposed to use bootstrapping, and Yuan and MacKinnon (2009) and Zhang, Wedel, and Pieters (2009) later proposed to use Bayesian estimation to establish the significance of the indirect effect. These techniques provide a confidence or credible interval (typically 95%), which need not be symmetric around the mean estimate of the indirect effect (Kisbu-Sakarya, MacKinnon, and Miočević 2014; Shrout and Bolger 2002). The indirect effect is deemed statistically significant if its confidence or credible interval does not contain zero. Assessing the significance of the indirect effect is facilitated by routines in statistical programs for bootstrapping (e.g., Hayes 2012; Preacher, Rucker, and Hayes 2007;



Shrout and Bolger 2002) and Bayesian estimation (Muthén, Muthén, and Asparouhov 2016). In the mediation review 94% of the analyses used the confidence interval from bootstrapping to determine whether the indirect effect was significantly different from zero (table 1). The others reported *p*-values or verbally summarized results. All three mediation models described here are forms of statistical mediation analysis because they rely on estimating the indirect treatment effect through an observed mediator.

### MEANINGFUL MEDIATION ANALYSIS

Using statistical mediation analysis to make causal inferences about indirect treatment effects rests on the heroic assumption that it transforms correlation into causation. That is, because of experimental manipulation and random assignment of participants to conditions, the statistical association between X and M (a-path), and between X and Y (c-path) can be interpreted causally. These are the strongest links. Yet, the b-path between the mediator and outcome is a (partial) correlation. It is the weakest link, and prone to various biases. What is more, bias in the b-path between mediator and outcome propagates to bias the conditional direct effect (c'-path), because the total treatment effect is the fixed sum of the indirect and conditional direct effects. Causal inferences from statistical mediation analysis become more plausible when the mediator-outcome correlation meets six conditions. Figure 2 and table 2 summarize the conditions, potential biases, and recommendations.

\*\*\* Insert figure 2 and table 2 \*\*\*

#### The Directionality Condition

The directionality condition specifies that the most plausible direction of influence is from the mediator to the outcome. Although it is crucial, this condition is seldom considered in

the mediation literature. When it is not met, the wrong direction of the causal chain is inferred or a causal chain is inferred where none exists.

By design, causal influence starts at the treatment and runs from there, respectively, to the mediator and to outcome. However, the causal direction between the mediator and outcome is undetermined. That is, on statistical grounds it is equally likely that (1) the mediator influences the outcome, (2) the outcome influences the mediator, or (3) the mediator and outcome are two correlated consequences of the treatment. The correlation between mediator and outcome might even result from a process where mediator and outcome simultaneously influence each other, and ignoring this would lead to simultaneity bias. The three ways to specify the correlation between mediator and outcome (1 to 3) produce three statistically equivalent models (MacCallum and Austin 2000; Williams 2012). Statistically equivalent models have the same implied covariance matrix, and thus the same global fit in terms of BIC and similar criteria. Therefore, the causal direction between mediator and outcome cannot be established on statistical grounds when only information about treatment, mediator, and outcome is available. With a single mediator and outcome there are (at least) three equivalent models. With two mediators and one outcome there are already twenty-seven equivalent models (25 recursive and 2 non-recursive), and all these models are statistically equally likely. The multiple mediation model in Figure 1 shows this. Not only are the links between the two mediators and the outcome correlational, but the d-link between the two mediators is so as well, and all these links are causally undetermined.

Although equivalent models have the same global statistical fit, they may differ in the size and significance of the indirect effect, and in local fit criteria such as the  $R^2$  for a specific presumed mediator or outcome. It is thus tempting to use such local criteria as evidence that the hypothesized mediation model is superior to a “reverse mediation model” in which the causal

direction is assumed to run from the outcome to the mediator. It is even tempting to “let the data speak”, by inspecting several equivalent models, and then selecting the model with the largest indirect effect as evidence that the “true” causal chain has been identified. These are not good ideas. The path between mediator and outcome is a (partial) correlation. It does not imply causation, and local fit criteria cannot serve as evidence for the plausibility of one causal process over others (Preacher and Kelley 2011, pp. 100-102; Roberts and Pashler 2000). As a case in point, Thoemmes (2015) found in a large simulation study that an inspect-and-select strategy in mediation analysis identified the known, true model in less than 30% of the cases for a basic mediation model where influence in each of the three paths could have two directions. An inspect-and-select strategy also invites data-driven rather than theory-driven analyses (Ioannidis 2005), and hypothesizing after the results are known (HARKing: Kerr 1998) which renders statistical testing meaningless.

To further illustrate the issue, let a treatment ( $X$ ) have a correlation of .50 with some outcome A, a correlation of .25 with another outcome B, and let outcomes A and B be correlated .50. Under the hypothesis that outcome A is the mediator and outcome B is the final outcome (model 1), the indirect effect ( $a \times b$ ) is  $.50 \times .50 = .25$ , the conditional direct effect ( $c'$ ) is .00, and the total treatment effect ( $c = a \times b + c'$ ) is .25. However, under the hypothesis that outcome B is the mediator and outcome A is the final outcome (model 2), the indirect effect ( $a \times b$ ) is  $.25 \times .40 = .10$ , the conditional direct effect ( $c'$ ) is .40, and the total treatment effect is ( $c = a \times b + c'$ ) is .50. And under the hypothesis that outcomes A and B are the final, correlated outcomes (model 3), there is no indirect effect. Then, the total treatment effect is the sum of the two direct effects, which is  $.50 + .25 = .75$ . Thus, model 1 has the largest indirect effect (.25), model 2 has the

largest conditional direct effect (.40), and model 3 has the largest total effect (.75). Yet, the three models are statistically equally likely, and a theoretical case for each might be made.

In fact, frequently “The substantive theory has a host of alternatives, some of which are interesting theoretically, some of which are not, and most of which nobody has thought of but could in a morning’s free-wheeling speculation” (Meehl 1990, p. 229). It is good advice to explicitly consider equivalent models and to provide evidence, other than statistical fit, for the superiority of the hypothesized model. In spite of this, 67-out-of-68 relevant studies published in top management journals focused on a single model without considering the presence of theoretically distinct but statistically equivalent models (Shook, Ketchen, Hult, and Kacmar 2004; see also MacCallum and Austin 2000). In the mediation review, statistically equivalent models were considered in only 8 cases (5% of 166). None of these pointed out the statistical equivalence of the models, and all used the statistical significance of the indirect effect to claim the superiority of the hypothesized model over the alternatives.

Kupor and Tormala (2015, Study 3) reported a serial mediation analysis with two mediators, and provided the data for re-analysis. Based on attitude and persuasion theory, the authors tested the hypothesis that interruption of a persuasive message (X) increases people’s curiosity (M1) which increases their favorable thoughts (M2) which improves behavioral intentions with respect to the topic of the message (Y). The re-analysis here examined statistically equivalent alternative models. Details are in appendix A. To keep the number of models manageable, only the causal direction between the two mediators was varied, as in figure 1. Model 1 is the serial mediation model hypothesized by the authors. Model 2 is an alternative serial mediation model. It specifies that the causal flow is from favorable thoughts (M2) to curiosity (M1). This model is consistent with the idea that subjective experience has a valence

and arousal component, and that more favorable thoughts (valence) increase curiosity (arousal) (e.g., Kuppens, Tuerlinckx, Russell, and Barrett 2013) which drives behavioral intentions. Model 3 is a parallel mediation model, which specifies two separate, but potentially correlated, pathways of the treatment effect respectively via curiosity (arousal) (M1) and via favorable thoughts (valence) (M2). In both serial mediation models, the serial indirect effect is significantly different from zero. In all three models, the two parallel indirect effects are also significantly different from zero. The serial mediation model that the authors proposed is firmly rooted in persuasion and attitude theory. Yet plausible alternative but statistically equivalent models were not considered. The re-analysis examined two of these.

A strong case that the proposed mediation model is more plausible than statistically equivalent models rests on logic, theory, and prior research findings (Meehl 1967, 1990). A strong case indicates explicitly why alternative models are less plausible. This may be straightforward to do in case, say, a treatment influences visual attention to various brands on a comparison website, which is followed by a choice for one of the brands. Then, it is more likely that attention predicts subsequent choice, rather than the other way around. However, making a strong case for the superior plausibility of the hypothesized mediation model vis-à-vis statistically equivalent alternatives is often more challenging. This is particularly so when all mediator and outcome measures are collected in a single experimental session, which is common. To illustrate, not only do people's appraisals of events influence their emotions which then influence their choices, but the emotions also activate the appraisals which influence choices, and the choices themselves can impact both the appraisals and the emotions (Lerner, Li, Valdesolo, and Kassam 2015). Such dynamics make it hard to establish, at a single point in time, the causal pathways between observed appraisals, emotions, and choices. The directionality

condition is even more critical in multiple mediation models. Then, the task to build a strong case for the superior plausibility of the hypothesized mediation model grows with the growing numbers of statistically equivalent alternative models.

If a strong case for a specific causal direction between mediator and outcome, or between mediators, cannot be made it is good advice to refrain from statistical mediation analysis and treat the presumed mediators and outcome as separate, correlated outcomes of the treatment (Bullock and Ha 2011, p. 517). This may be sufficiently interesting by itself.

Longitudinal mediation analysis is a promising approach to move closer to establishing causal direction (Cheong, MacKinnon, and Khoo 2003; Preacher 2015). Then, the mediator(s) and outcome(s) are measured repeatedly at different points in time after or while administering the treatment. This makes it possible to examine the growth trajectories of mediators and outcomes, and how these potentially influence each other over time. There do not seem to be published examples in the consumer behavior literature yet.

Experimental mediation analysis can establish causal direction between mediator and outcome. Then, a sequence of studies examines the influence of, respectively, the treatment on the observed mediator and outcome, the manipulated mediator on the outcome, and sometimes the manipulated outcome on the mediator (Smith 1992; Spencer, Zanna, and Fong 2005). Thus, each step in the causal chain after the treatment is manipulated and observed, which enables unbiased tracing of causal effects along the chain. Experimental mediation analysis requires valid and strong manipulation of the targeted mediator, and results in a loss of efficiency and systematic bias when such manipulations are unavailable (Bullock, Green, and Ha 2010). Tailored designs to implement experimental mediation analysis have been proposed (Imai, Tingley, and Yamamoto 2013). Lisjak, Bonezzi, Kim, and Rucker (2015) effectively used an

experimental mediation analysis to test the theory that compensatory consumption activates rumination, and that rumination in turn influences self-regulation activity.

### The Reliability Condition

The reliability condition specifies that measurement error in mediator and outcome is ignorable. When the condition is not met, the indirect treatment effect is *attenuated* towards zero. This attenuation bias increases the probability of failing to identify a true non-zero indirect effect (Type II error). Attenuation of the indirect effect also biases the conditional direct effect, but in various possible ways. When the true indirect ( $a \times b$ ) and conditional direct effect ( $c'$ ) have the same sign (complementary mediation; Zhao, Lynch, and Chen 2010), attenuation of the indirect effect leads to exaggeration of the conditional direct effect. Yet, when the true indirect and conditional direct effect have opposite signs (competitive mediation), attenuation of the indirect effect leads to attenuation of the conditional direct effect as well.

Cronbach's coefficient alpha is a reliability estimate for multi-item measures. It is considered satisfactory at .70, good at .80, and excellent at .90. An early meta-analysis of 832 studies in marketing and psychology found average reliabilities ranging from .70 for values and beliefs to .82 for job satisfaction, with an overall mean of .77 (Peterson 1994).

In the mediation review, reliability coefficients for multi-item measures were on average good, with a large range: average .84 for the mediator (range .51 to .97; average  $k$  items = 2.94) and .85 for the outcome (range .53 to .99;  $k$  items = 2.28). Still, even when reliability is good, just using the average scores of items for the mediator and outcome can lead to noticeably biased estimates of the indirect effect. That is, the observed correlation between mediator and outcome is a function of the true correlation and the reliabilities (Spearman 1904, p. 90):  $r_{YM}^{observed} =$

$r_{YM}^{true} \sqrt{(r_{MM}r_{YY})}$ , where  $r_{YM}^{true}$  is the “true” correlation,  $r_{YM}^{observed}$  is the “observed” correlation between mediator and outcome when the unreliability of their measures is not accounted for, and  $r_{MM}$  and  $r_{YY}$  are their reliabilities. If the true mediator-outcome correlation is .30 and the reliability of both measures is .80, the observed correlation is attenuated to .24. The biasing effects of measurement error exacerbate and rapidly become intractable in multiple mediation and other complex models (Cole and Preacher 2014).

The reliability condition can be met by improving the reliability of construct measures, and by estimating a structural equation model (SEM) to account for the remaining measurement error (Bagozzi 1977; Cole and Preacher 2014; Iacobucci, Saldanhi, and Deng 2007; VanderWeele, Valeri, and Ogburn 2012). Single item-measures of the mediator (43%), the outcome (64%) or both (34%) are common in consumer research, as the mediation review indicates. Single-item measures typically have lower reliability than multi-item measures. Importantly, structural equation modeling can also account for measurement error when mediator and/or outcome are measured with single items, if estimates of their reliability are available from previous studies, meta-analyses, or other sources (Anderson and Gerbing 1988; Westfall and Yarkoni 2016). For example, if a single-item measure of, say, the mediator has been part of a multi-item measure with known reliability, its reliability can be estimated using the Spearman-Brown prophecy formula as:  $r_{MM}^{single} = \frac{r_{MM}}{r_{MM} + k(1-r_{MM})}$ , with  $r_{MM}^{single}$  the estimated single-item reliability,  $r_{MM}$  the known multi-item reliability, and  $k$  the number of items in the multi-item measure, assuming all items are equally good. Using the results from the mediation review illustrates the approach. The average multi-item reliability was .84 for the mediator (average  $k = 4.00$  items, across 68 analyses reporting reliability of multi-item measures) and .86 for the outcome (average  $k = 4.88$  items, across 42 analyses reporting reliability of multi-item



measures). Estimated average single-item reliabilities in the mediation review are then, respectively, .57 for the mediator and .56 for the outcome (overall average .56). If no other information is available, these values might be useful as rough estimates of single-item reliability of mediator and outcomes measures. Other approaches to estimating single-item reliability are available (Bergkvist 2015; Wanous and Hudy 2001). Some analyses in the mediation review used a SEM program to estimate a traditional mediation model. None estimated a structural equation model to account for measurement error.

Kim and McGill (2011, Study 2) reported a moderated mediation analysis, measured the mediator and outcome with multiple items, and provided the raw data for re-analysis here. Details are in appendix B. The authors examined the effect of the interaction between perceived power (X1: high or low) and anthropomorphism (X2: low or high) on control perceptions (M; 4 items, alpha .69) and via those on risk perceptions about contracting skin cancer (Y; 3 items, alpha .91). The re-analysis compared a traditional model, with average scores and without control for measurement error, with a model that accounts for measurement error using SEM. Both analyses find evidence for the hypothesized moderated mediation effect ( $a_3 \times b$ ). Yet, when using a structural equation model the indirect effect was larger (-.20 versus -.11) and the conditional direct effect was smaller (-.13 versus -.17). Also 60% of the total effect was then mediated, as compared to 38% in the traditional model. One might thus argue that the traditional model is conservative and that accounting for measurement error will only result in larger indirect effects. However, the traditional model also overestimated the conditional direct effect. Moreover, obtaining unbiased estimates of indirect and direct effects may be more meaningful for theory and practice than establishing that estimates pass a statistical significance criterion (Gelman and Carlin 2014). A later section on reporting of effect sizes returns to this.

It is good practice to strive for high reliability of the mediator and outcome measures, and to account for the remaining unreliability by using a structural equation model (table 2).

### The Unconfoundedness Condition

The unconfoundedness condition specifies that variables which are omitted from the mediation model have ignorable effects on the association between mediator and outcome. Not meeting this condition leads to omitted variable bias (Imai, Keele, and Yamamoto 2010; Judd and Kenny 1981; VanderWeele, Valeri, and Ogburn 2012). Its implications depend on whether the omitted variables are pre-treatment or post-treatment, and on their correlations with the included variables.

Pre-treatment omitted variables are *not* influenced by the treatment themselves, and due to random assignment do *not* differ between conditions, but they influence both the mediator and outcome (Bullock, Green, and Ha 2010). For instance, age (U) may influence participants' desire for consistency (M1) and brand loyalty (Y), independent of a treatment (X) to stimulate the latter via the former. When such variables are omitted from the mediation model, the indirect effect of the treatment is biased, and because bias propagates the conditional direct effect is biased as well, and in the opposite direction. Specifically, the indirect effect ( $a \times b$ ) derived from the correlations between variables (Mauro 1990), with U denoting the omitted variable is:  $a \times b = r_{MX} \frac{r_{YM} - r_{YX}r_{MX} - r_{MU}r_{YU}}{1 - r_{MX}^2 - r_{MU}^2}$ . Omitted variable bias can exaggerate and attenuate the estimated indirect treatment effect. To illustrate this, let a treatment (X) have a correlation of .50 with the mediator (M), a correlation of .25 with the outcome (Y), and let mediator and outcome be correlated .50. In scenario 1, omitted variable U has an actual correlation of .50 with both mediator (M) and outcome (Y). Yet, because U is left out of the model, both correlations are

assumed to be zero. Then, the estimated indirect effect is .25 ( $a \times b = .50 \times .50$ ), while the actual indirect effect is only .125 ( $a \times b = .50 \times .25$ ). Omitted variable bias then exaggerates the indirect effect by a factor 2. Moreover, because bias propagates, the conditional direct effect is then falsely estimated to be zero ( $c = .25$ , and  $a \times b = .25$ ), although it is actually 50% of the total treatment effect ( $a \times b + c' = .125 + .125$ ). In scenario 2, the omitted variable  $U$  has an actual correlation of .50 with the mediator ( $M$ ) and of .10 with the outcome ( $Y$ ), which are not accounted for. Omitted variable bias then attenuates the indirect effect by 23% from the actual .325 ( $.50 \times .625$ ) to the estimated .25 ( $a \times b = .50 \times .50$ ). Such exaggeration and attenuation of indirect and conditional direct effects also increase the probability of Type I and Type II errors.

Post-treatment omitted variables are influenced by the treatment, and correlated with the outcome, and thus are omitted mediators (Bullock and Ha 2011; Imai, Keele, and Yamamoto 2010; Shrout and Bolger 2002). For instance, a treatment ( $X$ ) to encourage brand loyalty ( $Y$ ) via desire for consistency ( $M1$ ; included) may also stimulate satisfaction with the current brand ( $M2$ ; omitted) which can impact brand loyalty as well. Bias from omitted mediators might be large but is hard to predict, because it depends on the signs and sizes of the correlations between the treatment, the included and omitted mediators, and the outcome (Mauro 1990, p. 316, eq.2).

Confounding from omitted variables can be due to methodological similarities, such as when mediator and outcome are both measured with self-reports with similar item formats or response scales (number of responses, labels, fonts, and colors) or measured in close spatial or temporal proximity. This results in “common method bias.” For example, the correlation between two items was over 2 times larger when the items were measured consecutively rather than separated by six other items (reported in Podsakoff, MacKenzie, and Podsakoff 2012).

Confounding from omitted variables can also be due to substantive similarities between mediator and outcome. Conceptually related mediators and outcomes are more likely to share the same omitted causes. For instance, attitudes toward the ad and toward the advertised brand are probably associated with similar personality traits and current moods. More generally, the “crud factor” is likely to confound the mediator-outcome correlation, because in social science “everything correlates to some extent with everything” (Meehl 1990, p. 204). Such correlations arise because any measured psychological state or trait is some function of other states and traits. As a consequence, it is unlikely that correlations between measured variables, such as mediators and outcomes, are exactly zero (Lykken 1991). In the mediation review, only two analyses, both in Valsesia, Nunes, and Ordanini (2016), considered the possibility of omitted variable bias.

The unconfoundedness condition is in-the-limit untestable, because one can never be completely certain that the mediator-outcome correlation is unconfounded by omitted variables. But there are various prevention and coping strategies.

First, omitted variable bias can be reduced by measuring the mediator and outcome with different instruments, item types, or response scales, and by increasing the space and time interval between measurement of mediator and outcome (Podsakoff, MacKenzie, and Podsakoff 2011). For example, Kim and McGill (2011), Kupor and Tormala (2015) used a variety of measures and response scales for the mediators and outcomes which should reduce common method bias. Statistical procedures to correct for remaining bias are available (Bagozzi 2011; Richardson, Simmering, and Sturman 2009).

Second, adding potential confounders to the mediation model, based on theory of their potential influence on the mediator and outcome independent of the treatment, can make the unconfoundedness condition more plausible (Fiedler, Schott, and Meiser 2011; Pearl 2009;

Westfall and Yarkoni 2016). For instance, the re-analysis of Kim and McGill (2011) tested but found no evidence for potential confounding due to participants' gender and age (appendix B).

Third, sensitivity analyses can establish at which level of omitted variable bias the indirect effect would become indistinguishable from zero (Mauro 1990). Causal inferences are more plausible when the indirect treatment effect remains statistically significant even if mediator-outcome correlations due to potential omitted variables would be large. According to Imai, Keele, and Tingley (2010, p. 315): "... a mediation analysis is not complete without a sensitivity analysis." Routines for such analyses are available (Hicks and Tingley 2011; Muthén, Muthén, and Asparouhov 2016). The re-analysis of Kim and McGill (2011) here included a sensitivity analysis. It showed that if omitted variables account for a mediator-outcome correlation of  $-.13$  from the original correlation of  $-.38$  (remaining correlation  $-.25$ ), the indirect treatment effect would stop to be statistically significant. Although there are no guidelines yet, the obtained indirect effect in the original study appears rather sensitive to omitted variable bias.

Fourth, experimental mediation analysis with random assignment to manipulated mediators can ensure unconfoundedness of the path from mediator to outcome, see also the directionality condition. Other techniques to meeting the unconfoundedness condition, such as instrumental variable (IV) estimation and related techniques (Preacher 2015; Zhang, Wedel, and Pieters 2009), do not seem to have been applied yet in consumer research.

The first three conditions aim to obtain unbiased estimates of the indirect treatment effect, by preventing endogeneity bias (Bullock and Ha 2011; Zhang, Wedel, and Pieters 2009). Then, the most plausible direction of influence is from mediator to outcome (directionality condition), and bias due to errors-in-variables (reliability condition) and omitted variables

(unconfoundedness condition) are ignorable. The next three conditions aim to obtain meaningfully-sized estimates of indirect treatment effect.

### The Distinctiveness Condition

The distinctiveness condition specifies that mediator and outcome are distinct constructs. When measures of two constructs are empirically distinct, they express discriminant validity. Without discriminant validity, mediator and outcome are not empirically distinguishable, even though they might be conceptually. Then, the mediator and outcome measures are more properly treated as substitute measures of a single outcome. In view of its importance, it is surprising how rarely discriminant validity is treated in the mediation literature, as noted by Zhao, Lynch, and Chen (2010), and in causal inference more generally (Shook, Ketchen, Hult, and Kacmar 2004).

There are several criteria for discriminant validity (Voorhees, Brady, Calantone, and Ramirez 2016). A lenient criterion is that the “true” correlation between mediator and outcome, i.e., after correcting for measurement error, is less than one (Fornell and Larcker 1981) ( $r_{MY}^{true} < 1$ ). This criterion is met when a measurement model that fixes the true correlation between mediator and outcome to be one fits worse than a model that estimates the correlation freely. A strict criterion is that the true variance in the measures of, respectively, mediator and outcome is larger than the true variance that mediator and outcome share (Fornell and Larcker 1981) ( $r_{MM}, r_{YY} > r_{MY}^{true^2}$ ). This criterion is met when the reliability of the measures of mediator and outcome is larger than their squared true correlation. The average variance extracted (AVE) by mediator and outcome is a reliability estimate for multi-item measures of mediator and outcome.

The probability of establishing discriminant validity is higher when reliabilities of the mediator and outcome measures are higher, when true correlations between them are smaller,

and when sample sizes are larger. All analyses in the mediation review reported the sample size. But only 42% of all cases reported reliability estimates for the mediator measures, and only 27% did for the outcome measures. Merely 34% reported some estimate of the b-path between mediator and outcome. Together this makes it hard to judge overall discriminant validity between mediator and outcome measures in the mediation review. Three of the 166 cases did provide evidence of discriminant validity. As an example of such good practice, Valsesia, Nunes, and Ordanini (2016, Study 4,  $n = 280$ ) first established discriminant validity. They found that the average variance extracted by mediator (perceived creative authenticity of a new product: .58) and by outcome (product evaluation: .63) were higher than their shared variance (.39). Then the authors tested and found support for the hypothesis that the interaction between creative control (low-high) and investor trustworthiness (low-high) influenced the outcome via the mediator.

Developing and testing theories with precise definitions of mediator and outcome, and large conceptual leaps between them increases the probability that their measures express discriminant validity (Podsakoff, MacKenzie, and Podsakoff 2016). Measuring constructs with high reliability, using large sample sizes does so as well. Meaningful mediation analysis rests on discriminant validity between mediator and outcome. Recommendations are in table 2. The following sections return to the issue.

### The Power Condition

The power condition specifies that there is sufficient statistical power to identify the true relationships between treatment, mediator, and outcome, and their differences. Not meeting the power condition can have various biasing effects, including that indirect or conditional direct treatment effects are deemed not different from zero although they are (Type II error).

Statistical power is the probability of identifying a true non-null effect, with 80% as a common benchmark (Cohen 1962; Sawyer and Ball 1981). Then, in 80% of an infinitely large sequence of re-running a study on independent random samples from the population, the null hypothesis is correctly rejected at the chosen significance level, usually 5%. The statistical power of hypothesis testing in behavioral and health sciences is often low (Ioannides 2005; Maxwell 2004; Sedlmeier and Gigerenzer 1989). To illustrate, Button et al. (2013) found a median power of 21% across 730 neuroscience studies. Likewise, Cashen and Geiger (2004, p. 160) found that only 4 out of 53 articles they examined in management had sufficient statistical power to draw valid conclusions about their null-hypothesis tests. Statistically significant effects obtained in underpowered studies might be false positives due to random fluctuations or to exercising one's researcher degrees of freedom (Simmons, Nelson, and Simonsohn 2011).

Larger effect sizes and larger sample sizes, at the chosen significance level, increase the statistical power of finding true non-null effects. The commonly small-to-moderate effect and sample sizes in psychology and marketing contribute to low statistical power.

Correlations of .10, .30 and, .50 are deemed small, moderate, and large effect sizes (Sawyer and Ball 1981). A meta-analysis of 322 earlier meta-analyses in social psychology found an average correlation of .21 (Richard, Bond, and Stokes-Zoota 2003). A replication exercise of 100 studies published in psychology (Open Science Collaboration 2015) found an average correlation of .20. A meta-analysis of effect sizes in 94 meta-analyses in marketing found an average correlation of .27 (Eisend and Tarrahi 2014).

A review of psychology journals reported a 30-year stable average sample size of around 100 (Marszalek, Barber, Kohlbart, and Holmes 2011). The average sample size across 177 mediation analyses in psychology and marketing journals between 2011 and early 2014 was 151



(Peters 2017). At  $n = 183$ , the average sample size in the mediation review here is higher. This is partly due to the increased use of MTurk samples (73 cases:  $n = 215$ ,  $SD = 148$ ) rather than other, mostly student, samples (93 cases:  $n = 158$ ,  $SD = 93$ ; difference test  $p = .003$ ).

None of the analyses in the mediation review reported the statistical power of testing the hypothesized indirect treatment effect, or reported the sample size estimation procedure. Despite this, a high 99% of all 166 analyses reported having obtained evidence for the hypothesized indirect effect. Statistical power concerns the probability of finding a statistically significant indirect treatment effect, conditional on the mediator and outcome measures expressing discriminant validity. The re-analysis of Kim and McGill (2011) indicates that a sample size of about 500 is needed for the mediator to pass the strict discriminant validity criterion at 80% power, which is about six times larger than used. At that sample size, the conditional direct effect of the interaction variable on the outcome would also be significant at 80% power. The re-analysis of Kupor and Tormala (2015) indicates that the hypothesized serial mediation effect requires a sample size of about 450 to be significant at 80% power, which is about twice as large as used. Statistical power is a particularly important condition in mediation analysis, because it differs markedly between the indirect, the conditional direct, and the total effect.

Power Advantage of the Indirect Effect. Using Monte Carlo simulations, Kenny and Judd (2014) found that tests of the indirect effect ( $a \times b$ ) tend to have more power than tests of the total ( $c$ ) and conditional direct effect ( $c'$ ). This power advantage increases the probability that at a particular sample size the indirect effect is statistically significant, but the conditional direct and total effect are not. This may upward bias conclusions about the importance of the indirect effect, and prevent search for other or better mediators. To illustrate, when the indirect and total effect were both .09 ( $a = b = .30$ ;  $c = .09$ ), the required sample size for 80% power was 114 for

the indirect effect and 966 for the total effect. Likewise, when the sample size was sufficient to identify a significant indirect effect of .09 at 80% power, the power of the same-sized conditional direct effect of .09 was five times smaller at 16%. The power advantage of the indirect effect comes from splitting-up the total effect in several steps each with a larger effect and power than the total. The advantage is larger when the correlation between mediator and outcome is large (Kenny and Judd 2014; Pieters 2017). It increases in serial mediation models, in particular when the total treatment effect is small (O'Rourke and MacKinnon 2015, table 9).

It is good practice to use prospective power analysis when estimating the required sample size for the next mediation study, and to consider the hypothesized indirect and direct effects, as well as discriminant validity in such an analysis. Retrospective power analysis after data collection helps to assess the potential meaningfulness of the obtained effects. Retrospective power analysis "... often seems to be used as an alibi to explain away nonsignificant findings" (Gelman and Carlin 2014, p. 2). More productively it can be used to establish the power of studies with statistically significant findings, for meta-analyses (Button et al. 2013) or when re-analyzing specific studies, as done here. Various resources for such analyses are available (Fritz and MacKinnon 2007; Gelman and Carlin 2014; Muthén, Muthén, and Asparouhov 2016). It remains good advice that "investigators use larger samples than they customarily do" (Cohen's 1962, p. 153), certainly for mediation analysis. The section on Sweetspot analysis returns to this.

### The Mediation Condition

The mediation condition is the final condition. It specifies that the treatment effect on the outcome is transmitted via one or more mediators. It is plausible that the mediation condition is met when the first five conditions are met, and the hypothesized indirect treatment effect is statistically significant. The flowchart in figure 2 illustrates this. Most cases in the mediation

review emphasized the last part of the mediation condition only, that the indirect effect be statistically significant.

Meeting the first five conditions contributes to the validity of the inferred mediation effect. The first three conditions, directionality, reliability, and unconfoundedness, contribute to nomological validity. They increase the probability that the true direction, size, sign, and significance of the link between mediator and outcome are identified, and thereby the true indirect and conditional direct effects. The fourth condition, distinctiveness, provides evidence for discriminant validity, which relies on the reliability and unconfoundedness conditions. The fifth condition, statistical power, adds statistical conclusion validity.

## THE SWEETSPOT

Inferences about indirect treatment effects rely on the size of the mediator-outcome correlation. The correlation needs to be large enough to be significantly different from zero, at 80% power, as a lower limit. The correlation needs to be small enough to express discriminant validity at 80% power, as an upper limit. The “Sweetspot” is the region of statistically meaningful correlations between these lower and upper limits. Correlations outside the region provide insufficient statistical evidence for meaningful mediation.

Conditions 2 to 5 jointly define the Sweetspot. It requires a non-null mediation effect (condition 6), and discriminant validity of the measures of mediator and outcome (condition 4), and both of these at sufficient statistical power (condition 5). Discriminant validity relies on the reliability of measures of mediator and outcome (condition 2). Finally, the mediator-outcome correlation should be unbiased by omitted variables (condition 3).

I propose Sweetspot analysis, to follow up on Zhao, Lynch and Chen's (2010) call for more attention to discriminant validity in mediation analysis, and Edwards and Berry's (2010) recommendation in management to examine permissible ranges of correlations between constructs. Simulations were used to explore the size of the Sweetspot under various scenarios. The simulations systematically varied the "true" correlation between mediator and outcome, the reliability of their measures to assess discriminant validity, and the sample size. True mediator-outcome correlation was varied in steps of .01 from .10 to .95. Average reliability of the mediator and outcome measures was varied in five levels (.50, .60, .70, .80, and .90) assuming a single or multiple-item measure with known reliability. Sample size was varied in nine levels (50, 100, 150, 200, 250, 300, 400, 500, and 1000). The resulting 3,870 scenarios were each estimated with 1,000 replications, using the Mplus program (Muthén and Muthén 2015) and MplusAutomation (Halquist 2016). Statistical power was deemed sufficient when 80% out of the 1,000 estimates of a parameter were significant at  $p = .05$ . The analysis assumes that the unconfoundedness condition is met.

Figure 3 summarizes the results for the "observed" mediator-outcome correlations, because all cases in the mediation review relied on these. Such correlations are not corrected for measurement error. They result from correlating the raw score of single-item or the average of multi-item measures of the mediator and outcome. Figure 3 helps to gauge whether an observed correlation between mediator and outcome is likely to express discriminant validity, for a particular sample size and measurement reliability. The horizontal axis in the plots represents the sample size of a study. The vertical axis represents the size of the observed correlation between mediator and outcome.

The curves at the top of figure 3 indicate the upper limit for each of the reliability levels, at 80% statistical power. The curve at the bottom indicates the lower limit of the observed correlation. The Sweetspot starts to the right of the point where upper and lower limits cross, and becomes wider when sample sizes increase. A wider Sweetspot indicates a larger region in which an observed mediator-outcome correlation is likely to be statistically meaningful. Observed correlations *inside* the Sweetspot for the strict criterion (right plot) are most likely to be meaningful. Correlations *outside* the Sweetspot for the lenient criterion (left plot) are least likely to be meaningful. The vertical line in the plot indicates the starting point of the Sweetspot for a reliability of .70, as a reference. Correlations to the left of this starting point are statistically not meaningful: they are too small to be different from zero, and too large for discriminant validity of mediator and outcome. Bullet points (a) and (b) in the plots are introduced later.

\*\*\* Insert figure 3 \*\*\*

Before inspecting the Sweetspot, one might suppose that only very large correlations, say around .90 and larger, fail the discriminant validity criteria. Actually, remarkably small mediator-outcome correlations already fall outside the Sweetspot, in particular for modest sample sizes and reliabilities. To illustrate, at a sample size of 100 or less and with reliabilities of .70 or less, there is essentially no Sweetspot for the strict criterion. At this sample size, a correlation cannot be simultaneously significantly different from zero and meet the strict criterion for discriminant validity, at 80% power. And, only correlations between .27 and .38 (see left plot in figure 3) are inside the narrow Sweetspot for the lenient criterion.

At a sample size of 200, which is larger than 66% of the samples in the mediation review (average  $n = 183$ ), and reliability of .80 for mediator and outcome, observed correlations larger

than .56 and .40 already fail, respectively, the lenient and strict criterion. The maximum statistically meaningful correlation drops further when reliabilities are lower. This is pertinent given the large proportion of cases in the mediation review that relied on single-item measures, with average estimated reliabilities of around .56. Even at reliabilities of .70, mediator-outcome correlations of .50 with a sample size of 200 are outside the Sweetspot for both discriminant validity criteria. Bullet point (b) indicates this. Moving a bullet point horizontally to the right in the plots until it falls inside the Sweetspot for a particular reliability level indicates the required sample size to support discriminant validity. For an observed correlation of .50 with reliabilities of .80, a sample size of about 150 suffices to meet the lenient criterion, but a sample size of over 500 is required to meet the strict criterion, at 80% power.

The observed mediator-outcome correlation in Kim and McGill (2011) was -.38, with an average reliability of .79, and sample size of 84. Bullet point (a) in figure 3 represents this. It is just inside the narrow Sweetspot for the lenient but not for the strict criterion. With a sample size of 223 and reliabilities of over .83, the mediator-outcome correlations in Kupor and Tormala (2015) were mostly inside the Sweetspot for both criteria.

The Sweetspot analysis here is the first to explore the region of statistically meaningful mediator-outcome correlations, using common criteria for reliability, discriminant validity, and statistical significance, and power. It is not intended as an automatic tool to make pass-fail decisions based on fixed cut-offs about the meaningfulness of observed mediator-outcome correlations. It provides regions of meaningful correlations to plan for and to assess, under various reasonable scenarios. Figure 3 illustrates its application. Future work can explore other discriminant validity and reliability measures and statistical criteria, non-linear mediation relationships, discrete and count mediators and outcomes, and within-subjects data.

Prospectively, Sweetspot analysis can help to estimate the required sample sizes and reliabilities for meaningful regions of correlations in future studies, building on information from past studies. Retrospectively, Sweetspot analysis can help to assess the meaningfulness of observed correlations in past mediation studies for meta-analysis and literature review. The summary of simulations here provides a start. Table 2 has recommendations.

## MEANINGFUL COMMUNICATION OF MEDIATION

Comprehensive communication of mediation analysis results contributes to insight into the causal process of key interest, and to knowledge accumulation across analyses and studies. Four components of such communication are examined (table 3).

### Effect Decomposition

Mediation analysis decomposes the total treatment effect in conditional direct and indirect effects. It is thus informative to report estimates of *all* indirect and conditional direct treatment effects and separate paths, rather than of the indirect effect only. Doing so makes it possible to compare the indirect to the total and direct effects, and to gauge how much of the total effect is transferred by the mediator. That also indicates whether the mediator is empirically closer to the treatment, to the outcome, or midway between these (Shrout and Bolger 2002), and it is needed to establish discriminant validity. Furthermore, it helps to assess the accuracy of the analysis and reporting (Petrocelli, Clarkson, Whitmore, and Moon 2013).

Information about the uncertainty around point estimates of the effects and specific paths adds further insight. Uncertainty is reflected in standard errors, confidence intervals, and *p*-values. Because *p*-values are continuous measures of fit between data and model, with 0 as

complete misfit and 1 as complete fit, reporting exact  $p$ -values is more informative than reporting stars and 5% or other cut-offs (Greenland et al. 2016). It is good practice to provide point estimates of all effects and paths, and the uncertainty around these.

In the mediation review, 94% of the cases reported the 95% confidence interval of the indirect effect. However, only 60% also reported a point estimate of the indirect effect. Merely 37% reported the conditional direct effect, and 25% reported the indirect effect and conditional direct effect. Estimates of the b-path were reported in only 34% of the cases. Such reporting practices make it hard to gain insight into the causal processes of interest.

It is good practice to report all model estimates. And presenting these in a few tables or in a boxes-and-arrows diagram rather than dispersed in the text facilitates their interpretation and comparison. Schrift and Amar (2015) illustrate such good practice in reporting on a multiple mediation analysis. Soster, Gershoff, and Bearden (2014) do so on a moderated mediation analysis. See also tables A1.2 and A2.2 in the appendices. Table 3 provides recommendations.

## Effect Size

Both standardized and unstandardized coefficients of the various effects and paths in a mediation analysis are informative. Standardized coefficients facilitate comparisons of direct and indirect effects, and specific paths within and between analyses (MacKinnon 2008; Preacher and Kelley 2011). This is useful for later meta-analyses, for sample size estimation when designing new studies, and it contributes to theoretical precision (Edwards and Berry 2010; Meehl 1967). Although not without their issues, standardized coefficients are “... often the most useful coefficient for answering questions about the influence of one variable on another, with or without other variables in the equation” (Cohen, Cohen, Aiken and West 2002, p. 157). It is useful to report the partially standardized effects sizes in mediation analysis. One way to obtain



these is by standardizing the measures of mediator(s) and outcome, and then interpreting the unstandardized coefficients from the analysis (Gelman 2008; Kim and Ferree 1981; Muthén, Muthén, and Asparouhov 2016). When the treatment (X) is binary and dummy-coded (1,0), the indirect treatment effect can be interpreted as the mediated change in standard deviation units of the outcome. This can be compared between analyses and studies (MacKinnon 2008; Miočević, O'Rourke, MacKinnon, and Brown 2017).

Unstandardized coefficients on raw data retain the original scaling of measures, which can be useful. To illustrate, in research on the influence that limited time budgets have on the efficiency of shopping as mediated by planning practices, it is informative to retain the original scaling of mediator and outcome in minutes and time spent planning and shopping (Fernbach, Kan, and Lynch 2015). Such meaningful, non-arbitrary metrics are rare in consumer research. Constructs like values, perceptions, desires, attitudes, and intentions are typically measured on arbitrary scales (Blanton and Jaccard 2006). Semantic differential or Likert items to assess such constructs can be on 5-, 7-, or 9-point response scales, and the meaning of scores depends on the scales and their coding. Even when scales are meaningful, such as response time in milliseconds, these may become arbitrary when applied to measure something else, such as social preferences or normative conflict. Interpretational difficulty increases when mediator and outcome are measured on different arbitrary scales. For example, in Kim and McGill (2011, Study 2) mediator and outcome were measured on, respectively, 7-point and 9-point scales. In Kupor and Tormala (2015, Study 3), the first mediator was measured on 9-point scales, the second (thought favorability) on a scale from -1 to +1, and the outcome was standardized. The choice of metrics stems from various theoretical, methodological, and practical considerations beyond the current

scope. Yet, leaving arbitrary and varying metrics of mediators and outcomes in their raw form hinders interpretation, comparison, and accumulation of effect sizes.

There are other effect size measures in mediation analysis, none without their issues (Preacher and Kelly 2011). The percentage or proportion mediated (PM) is a simple measure of the extent to which the mediator transfers the total treatment effect:  $(100 \times ab) / (ab + c')$  (MacKinnon 2008; Shrout and Bolger 2002). It is also known as the “ratio of the indirect to the total effect.” The percentage mediated provides graded conclusions about the extent of mediation. It prevents dichotomous and potentially ill-founded conclusions that mediation is “full”, “complete,” “indirect-only,” or not. To illustrate, five articles in the mediation review (6 analyses) concluded that mediation was full, complete, or indirect-only, and had information to calculate the percentage mediated. The percentage mediated in these cases was on average 48% (range 27 to 96), which is less than half-full. The percentage mediated also comes with limitations (Miočević, O’Rourke, MacKinnon, and Brown 2017). First, it is a relative measure, and can refer to a very small mediated effect if the total treatment effect is small. Second, it requires fairly large sample sizes for stable estimates (500 and up). Third, it can be negative or larger than one in case of competitive mediation, when the conditional direct effect and the indirect effect, and/or several indirect effects have different signs. Using the absolute values of the effects ( $ab$  and  $c'$ ) in such cases, and declaring this might alleviate some of the concerns. Keeping these caveats in mind, reporting the percentage mediated together with the total treatment effect is often informative. In Kim and McGill (2011) using SEM, the total effect of the hypothesized interaction on perceived risk was  $-.33$ , and 60% of it was mediated by

perceived control. In Kupor and Tormala (2015), the total effect of the interruption on behavioral intentions was .43, and 5% of it was due to the hypothesized serial mediation process.

Few cases in the mediation review reported estimates of all effects and paths. Cases that reported point estimates typically did not report whether these were standardized or unstandardized, and provided little information to calculate standardized estimates. Only a single analysis reported the percentage mediated, and it did not report other information. Other overall indirect effect sizes were not reported. The re-analyses of Kim and McGill (2011) and Kupor and Tormala (2011) report partially standardized coefficients, and the percentage mediated. It is good practice to report (partially or fully) standardized coefficients and overall effect sizes, to add unstandardized coefficients when mediator and outcome measures are on meaningful scales, and to declare what is being reported, see table 3.

### Difference Testing

In addition to knowing whether various indirect and conditional direct effects each differ from zero, it can be theoretically or managerially relevant to know whether effects differ from each other. In multiple mediation analysis, theory might suggest the hypothesis that indirect treatment effects transferred via various mediators differ from each other. In moderated mediation analysis, the hypothesis is that the indirect effects of groups defined by the moderator differ from each other. In mediation analysis, more generally, theory might suggest that the indirect effect captures more of the total treatment effect than the conditional indirect effect does.

Such hypotheses about differences between effects can be tested. Yet rather than using explicit difference tests, there is a tendency to conclude that two effects differ significantly from each other, when one effect is significant ( $p < .05$ ) and the other is not ( $p > .05$ ). Recall, the tendency to conclude that mediation is complete, full or indirect-only if the indirect effect is

significantly differently from zero and the conditional direct effect is not. Such conclusions might not be justified, because “the difference between significant and insignificant is not itself necessarily significant” (Nieuwenhuis, Forstmann, and Wagenmakers 2011, p. 1107). Take for example (from Gelman and Stern 2006) the following two estimates and standard errors of  $.25 \pm .10$  and  $.10 \pm .10$ . The first estimate is significant at 1% ( $z = 2.5, p = .01$ ) while the second estimate is not ( $z = 1.0, p = .32$ ). However, the difference between the two estimates is not statistically significant:  $z = 1.07, p = .28$ . Therefore, “... one should look at the statistical significance of the difference rather than the difference between their significance levels” (Gelman and Stern 2006, p. 329). Despite this, about half of 157 articles published in neuroscience claimed evidence for an interaction effect without reporting an explicit test for it (Nieuwenhuis, Forstmann, and Wagenmakers 2011).

Without evidence from an explicit difference test, conclusions about significant differences might still be justified when the available 95% confidence intervals are correctly interpreted. Yet, a common slip lurks here. When two 95% confidence intervals do *not* overlap each other, their difference is significant at the 5% level. However, when two 95% confidence intervals *do* overlap each other, their difference may still be significant at the 5% level, which is commonly missed (Belia, Fidler, Williams, and Cumming 2005). Non-overlap of confidence intervals is thus a conservative difference test (Schenker and Gentleman 2001). Explicit difference tests avoid the ambiguities in comparing confidence intervals. To prevent statistical testing rituals, such difference tests require hypotheses derived from theory.

Still, the analyses in the mediation review typically did not test differences between indirect effects in multiple mediation, and between the indirect effect(s) and the conditional direct effect. Also 38% (31 of 82) of the relevant analyses did not directly test for moderated

mediation, despite it being the central hypothesis. Then, support for moderated mediation was reported when the confidence interval of one group or condition did not overlap zero while the other did. In 14 of these cases, the two confidence intervals did not overlap each other, indicating a significant difference. However, in the remaining 17 cases, the average proportion overlap of the confidence intervals was a sizeable 20%. Table 3 summarizes the recommendations.

### Data Sharing

Sharing data from mediation analyses is a form of communication. It enables replication of findings, new analyses by others, quantitative literature reviews and meta-analyses. It may encourage a culture of openness and accountability, inspire knowledge synthesis, and generalizability testing. Data can be shared in raw or summary form.

Authors who submit to the *Journal of Consumer Research* commit to making the raw data and material available, upon request. Among the 86 articles in the mediation review, two author teams already did so upfront. Smith, Newman, and Dhar (2016) made their raw data available via the second-author's faculty pages at the university website. Jhang and Lynch (2015) provided open access to their raw data via a public, open-access repository (datadryad.org). Various public, open-access repositories offer permanent URLs, such as datadryad, Open Science Framework, Figshare, Github<sup>1</sup>, which is desirable.

Summary statistics data (SSD) are a compact, aggregate form of the raw data that can be readily included in reports. Because such data do not reveal individual responses, participant confidentiality is not violated, and proprietary data issues may play a lesser role. Summary statistic data typically include sample size, means, standard deviations, and correlations or

---

<sup>1</sup> [http://www.psychologicalscience.org/observer/finding-a-home-for-your-science#.WT\\_l-JeweAw](http://www.psychologicalscience.org/observer/finding-a-home-for-your-science#.WT_l-JeweAw), last accessed July 2017.

covariances of all variables, and reliabilities of multi-item measures. Path models or SEM can use such data for mediation analysis. They do not allow procedures that require raw data, like bootstrapping and Bayesian estimation, or data splits not conceived in the original study.

The re-analyses of Kim and McGill (2011) and Kupor and Tormala (2015) used the raw data shared by the authors, and report summary statistics data (tables A1.1 and A2.1). Further summary statistics data and input files are available online (). Reporting summary statistics data is common in strategic marketing research (e.g., Kirca, Jayachandran, and Bearden 2005). None of the analyses in the mediation review did. The informativeness and compactness of summary statistics data makes reporting them in manuscripts a desired default.

## CONCLUSION

This tutorial described six conditions to permit plausible causal inferences from statistical mediation analysis: directionality, reliability, unconfoundedness, distinctiveness, power, and mediation. Failure to meet the conditions leads to biases that rapidly become intractable, in particular in multiple mediation and other complex models, and when multiple conditions are not met. I proposed a framework that integrates these conditions, summarized in figure 2, and offered recommendations to meet and cope with them, summarized in table 2. Sweetspot analysis was proposed to explore the region of statistically meaningful mediator-outcome correlations. When all conditions are met, a statistically significant indirect treatment effect can be plausibly interpreted causally. Although all six conditions are important, the directionality and distinctiveness conditions in particular deserve more attention than so far. Without discriminant validity between mediator(s) and outcome, there can be no mediation. Without strong evidence for the hypothesized direction of influence from mediator to outcome, causal inferences from

mediation analysis are compromised. Then, Occam's razor favors a parsimonious model that estimates the direct treatment effect on two alternative measures of a single outcome.

To improve knowledge accumulation about mediation processes, I also described four components of comprehensive communication of mediation analysis results, and offered recommendations to improve current reporting practice, summarized in table 3. Jointly this can guide researchers when planning future mediation studies, when analyzing and appraising their own studies, and those of others before and after publication.

### Conditions and Communication

Meeting all conditions is challenging, and following the recommendations is likely to increase the length of reports. The decision to place some additional material in an appendix depends on the importance of the mediation analysis to the central questions that the research seeks to answer, the importance of the specific conditions, the results, and editorial policies.

It is not very useful for model selection or support to report the overall fit and parameter estimates of various statistically equivalent models (as I did in appendix A). These results were reported here only to point out that statistically equivalent models can lead to very different conclusions. Instead, it is paramount to make a strong case that the hypothesized model is more plausible than statistically equivalent models, using logical and theoretical arguments, and findings from prior research. This might add to the size of the theory section.

Further, stating in the main report that "The reliability, unconfoundedness, distinctiveness, and power conditions were met, see supporting materials (appendix X)" with the appropriate evidence may often suffice. Following the recommendations about comprehensive

communication will add one or two tables and some explanation in the main text. A table with summary statistics data can be in an appendix.

### Future Work

Other conditions for causal inference from mediation analysis (Pearl 2009) could have received more attention. For instance, the homogeneity condition specifies that average treatment effect holds similarly for all participants, which may be more an exception than the rule (Imai, Keele, and Yamamoto 2010; Valeri and VanderWeele 2013). Moderated mediation models, used by 49% of the cases in the mediation review, go a long way to meeting this condition for subgroups defined by manipulated or measured moderators, though it is hard to rule out all sources of heterogeneity. The homogeneity condition could have been treated more extensively.

The Stable Unit Treatment Value Assumption (SUTVA) could have been treated in detail. It specifies that knowledge of the treatment status of others does not influence participants' responses to their assigned condition (Ten Have and Joffe 2010). This condition may be hard to satisfy in field experiments or with crowd-sourced samples where assignment conditions may be or become known over time (Paolacci and Chandler 2014). Further research on the implications of data collection methods on the validity of theory testing is warranted.

The consumer behavior literature can gain from an increased focus on effect size measures, rather than on whether or not effects are significantly different from zero at the 5% level. Meta analyses rely on effect sizes. Prospective power analysis and sample size estimation also rely on them, and these are progressively called for by granting organizations and journals. Without a store of reasonable domain-specific effect sizes, sample size estimation has to fall



back on published “T-shirt” effect sizes (small, medium, and large), which may fit all except the needs of any specific study at hand.

A focus on effect size estimates can also contribute to more precise, quantitative theories and predictions, and move the discipline beyond the lamented null-hypothesis significance testing (Gelman and Carlin 2014; Meehl 1990). Correlations between observables in the social sciences are rarely exactly zero (Lykken 1991; Meehl 1967, 1990). The operation of the “crud factor” virtually ensures that with increased sample size and measurement reliability, the probability of rejecting a null hypothesis of a zero effect and of accepting any alternative hypothesis, even if it is false, increases. One way forward is to formulate hypotheses about reasonable effect sizes to expect (Edwards and Berry 2010). With increasing sample sizes and measurement reliabilities, the probability of accepting quantitative hypotheses increases, if these are correct. Such hypotheses will commonly be ranges rather than point estimates (Edwards and Berry 2010). In the words of the psychologist Lykken (1991, p. 33): “... we ought to be able to squeeze out of our theories something more than merely the prediction that A and B are positively correlated. If we took our theories seriously and made the effort, we should be able to make rough estimates of parameters sufficient to say, e.g., that the correlation ought to be greater than .40 but not higher than .80.” This is a worthy ambition for consumer research too.

In closing, consumer researchers are encouraged to meet all six conditions when conducting statistical mediation analysis, and to communicate the results comprehensively. This might result in fewer published mediation studies, and fewer mediation studies that support the hypothesized process. Meeting all six conditions is hard, but trying hard to meet them is likely to make causal inferences from statistical mediation analysis more plausible. A first step is to communicate the results of mediation analyses comprehensively.

## APPENDIX A

### RE-ANALYSIS KUPOR AND TORMALA (2015, STUDY 3)

Kupor and Tormala (2015, Study 3) conducted a serial mediation analysis on the influence that interruption of a persuasive message (X) has on behavioral intentions with respect to the topic of the message, as mediated first by curiosity (M1) and then by favorable thoughts (M2). A sample of 250 MTurk workers was randomly assigned to a treatment or control condition, and then responded to a set of questions. All participants saw a brief video on the health benefits of drinking coffee. In the treatment, and not in the control condition, the video was briefly interrupted for loading. Immediately after the video (X), participants indicated their curiosity (M1) on four nine-point items, from “not at all” to “completely” ( $\alpha = .93$ ). Then, they indicated their behavioral intentions (Y) to drink coffee and recommend it to others on three seven-point items, from “extremely unlikely” to “extremely likely”, and indicated the amount they were willing to pay for one cup of coffee. These four items were standardized and then averaged into the behavioral intention measure ( $\alpha = .83$ ). After that, participants indicated the favorability of each of their thoughts while watching the video. This was converted into a thought favorability index, which ranges from -1 in case all thoughts are unfavorable to 1 in case all thoughts are favorable. The 223 participants who reported at least one thought were included in the mediation analysis. A serial mediation analysis supports the theory that interruption of the video (X) raised curiosity (M1) which raised thought favorability (M2) which raised behavioral intentions (Y). Summary statistics data (SSD) are in Table A1.1.

The re-analysis focuses on the distinctiveness, directionality, and statistical power conditions. Prior to the re-analysis, measures for curiosity (9-point, ranging 1 to 9), and thought favorability (ranging -1 to 1) were standardized to place them on the same response scale; behavioral intentions had already been standardized by the authors. Treatment and control conditions were dummy coded, respectively, as 1 and 0.

#### Re-analysis

Distinctiveness. The re-analysis supports discriminant validity for the two mediators and the outcome vis-à-vis each other. Averaged data were available for the mediator and outcome measures, and the reliability of the measures of M1 and Y. Reliability of thought favorability (M2) is unavailable, since it is the net valence of the thoughts that participants generated.

Discriminant validity for this single-item measure of M2 was assessed under four scenarios, namely that its reliability was .93 (the same as M1), .83 (as Y), .60 or .50. The two scenarios with single-item reliabilities of .50 or .60 are realistic. If each separate item in a multi-item scale has a reliability of .50, the overall reliability would be .75 for a three-item scale and .80 for a four-item scale. Two criteria for discriminant validity were used. Lenient discriminant validity tested whether each of the three corrected (“true”) correlations between mediators and outcome were less than one (unity). Strict discriminant validity tested whether each of the three corrected (“true”) squared correlations ( $R^2$ ) between mediators and outcome was less than the reliability of the respective measures. Reliability is the true variance extracted by a construct from its measures, and strict discriminant validity requires that it is higher than the variance it shares with each of the other constructs (Fornell and Larcker 1981). For most scenarios, the two mediators and the outcome met both discriminant validity criteria. Only if the reliability of thought favorability (M2) would be .50, this mediator would fail the strict criterion vis-à-vis the outcome.

**TABLE A1.1**  
SUMMARY STATISTICS DATA (SSD): KUPOR AND TORMALA (2015, STUDY 3)

Label	Variables	Mean	SD	Correlations			
				X	M1	M2	Y
X	Interruption	.510	.501	--			
M1	Curiosity	6.583	2.181	.161	(.93)		
M2	Thought favorability	.594	.357	.159	.207	--	
Y	Behavioral intentions	.025	.796	.213	.356	.415	(.83)

NOTE.— $n = 223$ . Interruption (X) coded as 1 = yes ( $n = 114$ ), 0 = no ( $n = 109$ ). Curiosity is average of four 9-point items (1 to 9). Thought favorability is from -1 (all negative) to 1 (all positive). Behavioral intentions is average of four standardized items. Cronbach’s alpha of multi-item measures on the diagonal.

Directionality. Three equivalent multiple mediation models were estimated. In the first serial mediation model, which was hypothesized by Kupor and Tormala (2015), X first influences M1 which then leads to M2 and then to Y (d2-path in model 2 of figure 1). An alternative serial mediation model specifies that X first influences M2 which then leads to M1 and then to Y (d3-path). Finally, a parallel mediation model specifies that X influences M1 and M2 which then lead to Y. In this model, the residuals of the two mediators are allowed to correlate (undirected path, d1). All three models contain

the two parallel mediation pathways from X via M1 to Y, and from X via M2 to Y as defaults. Recall that the mediators and outcome were standardized prior to analysis. Table A1.2 gives unstandardized estimates,  $p$ -values and 95% confidence intervals, based on 15,000 bootstrapped samples.

All three models have the same global fit (BIC is 1876; model  $\chi^2(0) = 0$ ). Also, all three models show parallel mediation from X to M1 and to Y ( $a1 \times b1$ ), and from X to M2 and Y ( $a2 \times b2$ ), with the CI<sub>95</sub> not including zero. The two parallel mediation paths are equally strong in all three models. In addition, model 1 finds serial mediation from M1 to M2, and model 2 from M2 to M1. In all three models about 46% of the total treatment effect is mediated, and about 54% is direct ( $c'$ ). About 4-5% of the total treatment effect is mediated by a serial mediation pathway (model 1: .048, CI<sub>95</sub> [.009, .160], and model 2: .037, CI<sub>95</sub> [.005, .138]). In all three models, the total indirect treatment effect (.194) is smaller than the conditional direct effect (.233), which suggests that other mediators yet to be discovered may play a role.

Statistical power. Monte Carlo analyses (1,000 replications) were performed to examine the statistical power of obtaining the serial mediation effect. At the study's sample of 223, the hypothesized serial mediation effect ( $a1 \times d2 \times b2$ ) is statistically significantly different from zero (estimate .020, CI<sub>95</sub> [.004, .055]). The statistical power of obtaining this effect at  $n = 223$  is 23% which is lower than the recommended 80% power. To obtain such a statistical power, a sample size of about 450 would be needed in this study, which is about twice the current size. At that size, the conditional direct effect of the treatment (X) is also significant at 80% power.

Taken together, the re-analysis provides evidence for discriminant validity of the mediators and outcome on the lenient and strict criteria. The hypothesized serial mediation effect accounts for about 5% of the total treatment effect that interruption has on behavioral intentions, which is .426 (CI<sub>95</sub> [.165, .685]). A sample twice the current size is required to have 80% power of this hypothesized effect. All indirect effects together account for about 46% of the total treatment effect. This total indirect effect propagates mostly via two parallel pathways, curiosity (M1: 20% of total) and favorable thoughts (M2: 21% of total). The re-analysis also indicates that a (statistically equivalent) alternative serial mediation model and a parallel mediation model lead to different inferences about the causal process. Theoretical evidence provided by Kupor and Tormala (2015) favors the hypothesized serial mediation model.

**TABLE A1.2**  
**MEDIATION ANALYSIS: KUPOR AND TORMALA (2015, STUDY 3)**

Predictors	Criterion	Path	Hypothesized Model 1: Serial Mediation M1 to M2					Equivalent Model 2: Serial Mediation M2 to M1					Equivalent Model 3: Parallel Mediation M1 with M2				
			Est.	SE	p	CI <sub>95</sub>		Est.	SE	p	CI <sub>95</sub>		Est.	SE	p	CI <sub>95</sub>	
						LL	UL				LL	UL				LL	UL
X: Interruption	M1: Curiosity	a1	.321	.132	.015	.056	.585	.262	.134	.051	-.001	.526	.321	.134	.016	.056	.585
M2: Fav. thoughts		d3	--	--	--	--	--	.186	.070	.007	.049	.324	--	--	--	--	--
R <sup>2</sup>		--	.026	.022	.236	.001	.081	.060	.032	.065	.012	.126	.026	.022	.236	.001	.081
X: Interruption	M2: Fav. thou.	a2	.257	.131	.050	-.005	.513	.317	.132	.017	.056	.575	.317	.132	.017	.056	.575
M1: Curiosity		d2	.186	.069	.007	.048	.324	--	--	--	--	--	--	--	--	--	--
R <sup>2</sup>		--	.059	.033	.072	.010	.126	.025	.021	.240	.012	.126	.025	.021	.240	.001	.082
Correlation M1, M2		d1	--	--	--	--	--	--	--	--	--	--	.182	.069	.008	.052	.321
X: Interruption	Y: Intentions	c'	.233	.125	.063	-.009	.482	.233	.125	.063	-.009	.482	.233	.125	.063	-.009	.482
M1: Curiosity		b1	.267	.064	<.001	.140	.399	.267	.067	<.001	.140	.399	.267	.067	<.001	.140	.399
M2: Fav. thoughts		b2	.342	.063	<.001	.215	.468	.342	.064	<.001	.215	.468	.342	.064	<.001	.215	.468
R <sup>2</sup>		--	.261	.056	<.001	.151	.367	.261	.056	<.001	.151	.367	.261	.056	<.001	.151	.367
<i>Indirect effects:</i>																	
X to M1 to Y		a1 x b1	.086	.041	.036	.020	.187	.070	.038	.068	.007	.163	.086	.041	.036	.021	.187
X to M2 to Y		a2 x b2	.088	.049	.075	.006	.204	.108	.051	.034	.023	.228	.108	.051	.034	.023	.228
X to M1 to M2 to Y		a1 x d2 x b2	.020	.012	.096	.004	.055	--	--	--	--	--	--	--	--	--	--
X to M2 to M1 to Y		a2 x d3 x b1	--	--	--	--	--	.016	.011	.166	.002	.050	--	--	--	--	--
Total indirect of X		sum all a x b	.194	.071	.006	.071	.351	.194	.071	.006	.071	.351	.194	.071	.006	.071	.351
Total effect of X:		sum all a x b + c'	.426	.133	.001	.165	.685	.426	.133	.001	.165	.685	.426	.133	.001	.165	.685

NOTE.— $n = 223$ , with 15,000 bootstrapped samples. Interruption (X) coded as 1 = yes, 0 = no, and all other variables standardized before the analysis. Unstandardized coefficients (Est.) and their standard error (SE) shown. LL is lower level, and UL is upper level of 95% Confidence Interval.

**APPENDIX B**  
**RE-ANALYSIS KIM AND MCGILL (2011, STUDY 2)**

Kim and McGill (2011, Study 2) examined whether control perception (M) mediates the effect that the interaction (X1X2) between power (X1: high or low) and anthropomorphism (X2: low or high) has on risk perception about contracting skin cancer (Y). A sample of 84 students participated. Power perception was manipulated by having participants describe a personal experience in which they felt powerful (or powerless). For the anthropomorphism manipulation participants read a message about skin cancer that portrayed it as having humanlike intentions to hurt people (yes) or not (no). Control perception was measured with four 7-point “disagree” to “agree” items, and risk perception was measured with three 9-point “not at all” to “very much” items. The authors hypothesized that high power and low anthropomorphism, and low power and high anthropomorphism (the interaction: X1X2) would lead to the highest risk perceptions due to promoting the lowest control perceptions, and that the two main effects would have no effect on the outcome and mediator. This implies a mediated moderation model.

**TABLE A2.1**  
**SUMMARY STATISTICS DATA (SSD): KIM and MCGILL (2011, STUDY 2)**

Label	Variable	Mean	SD	Correlations (reliability)				
				X1	X2	X1X2	M	Y
X1	Power	.143	.996	-				
X2	Anthropomorphism	.000	1.006	-.048	-			
X1X2	Interaction X1 and X2	-.048	1.005	.007	.143	-		
M	Control perception	4.074	1.117	-.029	-.003	.322	(.69)	
Y	Risk	6.837	1.634	.145	-.037	-.271	-.377	(.91)

NOTE.— $n = 84$ . X1, X2, X1X2 are coded -1 (high) and +1 (low). Cell sizes: X1 high and X2 high = 17, X1 high and X2 low = 19, X1 low and X2 high = 25, X1 low and X2 low = 23. Control (M) is average of four 7-point items (range 1-7). Risk (M) is average of three 9-point items (range 1-9). Composite reliabilities in the diagonal.

The original moderated mediation analysis used the average scores of four mediator items and three outcome items. Our re-analysis examines the reliability, unconfoundedness, distinctiveness, and statistical power conditions. It differs slightly from the original, because the (non-significant) interaction between X1 and M on the outcome Y is not included here. Prior to the re-analysis, measures of the mediator (7-point scale) and outcome (9-point scale) were

standardized to place them on the same response scale. Main and interaction effects were effect-coded as, respectively, Power (X1) coded -1 = high, +1 = low, and Anthropomorphism (X2) coded -1 = high, +1 = low, and X1X2 as -1 and +1. Table A2.1 has summary statistics. The Mplus program was used in all analyses (Muthén and Muthén 2015).

### Re-Analysis

Reliability. Two models examined effect of measurement error. Model 1 is a traditional mediation analysis with average scores of mediator and outcome as in Kim and McGill (2011). Model 2 accounts for measurement error in mediator and outcome by using SEM. For both models bootstrapping (15,000 samples) was used. Table A2.2 summarizes the results.

Both models provide evidence for mediated moderation, with some notable differences. As Kim and McGill (2011) hypothesized, only the interaction variable (X1X2) has an indirect treatment effect and its conditional direct effect is not statistically significant. The mediated moderation effect is not statistically significant when assuming a Normal, symmetric distribution:  $a_3 \times b = -.105, p = .080$  in model 1, and  $-.201, p = .205$  in model 2, but it is when using bootstrapping: CI<sub>95</sub> [-.263, -.023] in model 1 and CI<sub>95</sub> [-.616, -.029] in model 2. When accounting for measurement error (model 2), the variance accounted for increases from 11% to 18% in the mediator, and from 19% to 24% in the outcome. Using a path model (model 1), the total treatment effect of the interaction variable is -.274 of which 38% is mediated, whereas these results are, respectively, -.334 and 60% when accounting for measurement error (model 2).

Also, the percentage mediated due to the interaction variable increases from 38% in model 1 to 60% in model 2. Model 2 estimates the indirect effect more accurately, with stronger evidence for mediation. Still, in both models the indirect effect of the interaction variable ( $a_3 \times b$ ) is not significantly different from its conditional direct effect ( $c'3$ ) (table A2.2).

Unconfoundedness. Two analyses examined potential omitted variable bias. First, model 2 was re-estimated with participants' age and gender as covariates of control (M) and risk perception (Y). Covariate selection is based on research that females perceive higher health risks (Harris, Jenkins, and Glaser 2006), and that risk perceptions about health-related behaviors increase and potentially control perceptions decrease with age (Bonem, Ellsworth, and Gonzalez 2015). Such effects might confound the relationship between control (M) and risk perceptions (Y). The covariates were insignificant ( $ps > .27$ ), and the mediated moderation effect remained

essentially the same ( $a_3 \times b = -.218, p = .254, CI_{95} [-.739, -.021]$ ), which is re-assuring. Second, a sensitivity analysis (Imai et al. 2010; Muthén, Muthén, and Asparouhov 2016) examined at which level of confounding the moderated mediation effect would become indistinguishable from zero. This would occur when omitted variables account for a correlation of  $-.13$  between M and Y ( $a_3 \times b = -.180, p = .204, CI_{95} [-.473, .008]$ ), which is only one-third of the observed correlation ( $-.38$ ). The mediation effect appears fairly sensitive to omitted variable bias.

Distinctiveness. Discriminant validity was assessed without the age and gender covariates based on their insignificant effect. Mediator and outcome meet the lenient discriminant validity criterion: A one-factor model (all seven measures on a single factor) fitted worse than a two-factor model (with the respective measures on separate factors): difference  $\chi^2(1) = 38.04, p < .001$ . The mediator did not pass a strict test (Fornell and Larcker 1981). At the sample size of 84, its average variance extracted (AVE) did not differ significantly from the variance it shares with the outcome: 37% versus 22% ( $p = .45$ , for the difference). The outcome passed the strict criterion: AVE versus variance shared was 76% versus 22% ( $p < .001$ , for the difference).

Statistical Power. A Monte Carlo analysis examined the statistical power of the mediation results, using 80% as cut-off criterion. For the analysis, 1,000 replication samples were generated, using the sample estimates of the SEM as population estimates. The percentage of parameters significant at  $p = .05$  across replications indicates their statistical power. At the sample size of 84, the statistical power of the mediated moderation effect is 73%. Statistical power for strong discriminant validity is 99% for the outcome, but only 33% for the mediator. A sample size of about 500 is needed to obtain strong discriminant validity for the mediator. At that sample size the conditional direct effect of power (X1) and its interaction with anthropomorphism (X1X2) on risk perception (Y) are also significant at 80% power.

Taken together, the re-analysis indicates that the control perceptions (M) mediate the effect that the interaction between power (X1) and anthropomorphism (X2) has on risk perception about skin cancer (Y) is mediated by control perception (M), and at sufficient statistical power in this sample. It further demonstrates that an about six times larger sample size (about  $n = 500$ ) is required for the mediator to pass a strict criterion of discriminant validity. At that sample size, the conditional direct effects of the interaction variable is also statistically significant at 80% power. The mediated effect is fairly sensitive to potential omitted variable bias. After controlling for measurement error about 60% of the total treatment effect is mediated.



**TABLE A2.2**  
**MEDIATION ANALYSIS: KIM AND MCGILL (2011, STUDY 2)**

Predictors	Criterion	Path	Model 1 Observed mediator and outcome (path)					Model 2 Latent mediator and outcome (SEM)				
			Est.	SE	p	CI <sub>95</sub>		Est.	SE	p	CI <sub>95</sub>	
						LL	UL				LL	UL
X1: Power	M: Control	a1	-.034	.104	.745	-.237	.171	-.063	.139	.650	-.319	.225
X2: Anthropo.		a2	-.051	.103	.617	-.257	.146	-.093	.139	.501	-.378	.167
X1X2		a3	.330	.103	.001	.129	.532	.465	.169	.006	.139	.817
R <sup>2</sup>		--	.108	.062	.081	.015	.212	.177	.095	.062	.018	.359
X1: Power	Y: Risk	c'1	.138	.107	.197	-.070	.350	.142	.146	.329	-.141	.433
X2: Anthropo.		c'2	-.007	.107	.949	-.219	.200	-.019	.145	.897	-.328	.242
X1X2		c'3	-.169	.120	.129	-.384	.051	-.132	.186	.477	-.438	.293
M: Control		b	-.318	.111	.008	-.576	-.109	-.432	.205	.035	-.900	-.109
R <sup>2</sup>		--	.186	.074	.012	.048	.304	.244	.104	.018	.068	.423
Indirect effects:												
X1 to M to Y		a1 x b	.011	.035	.759	-.059	.083	.027	.071	.700	-.096	.186
X2 to M to Y		a2 x b	.016	.034	.634	-.047	.095	.040	.069	.581	-.060	.236
X1X2 to M to Y		a3 x b	-.105	.060	.080	-.263	-.023	-.201	.157	.200	-.616	-.029
Total effect of X1X2:		a3 x b + c'3	-.274	.109	.012	-.488	-.063	-.334	.144	.021	-.611	-.053
Difference indirect and direct effect:		a3 x b - c'3	.064	.142	.653	-.240	.315	-.069	.313	.826	-.870	.343

NOTE.— $n = 84$ , with 15,000 bootstrapped samples. Measures of mediator (M) and outcome (Y) were standardized prior to analysis. Power (X1) coded as -1 = high, +1 = low, Anthropomorphism (X2) as -1 = high, +1 = low, and X1X2 as -1 and +1. Unstandardized coefficients (Est.) and their standard error (SE) shown. In model 2, the variances of latent mediator and outcome were fixed to 1 and loadings freely estimated for identification.  $R^2$  is variance accounted for in the criterion. *LL* is lower level, and *UL* is upper level of 95% Confidence Interval.

## REFERENCES

- Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3): 411–423.
- Bagozzi, Richard P. (1977), "Structural Equation Models in Experimental Research," *Journal of Marketing Research*, 14, 209-226.
- Bagozzi, Richard P. (2011), "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly*, 35(2), 261-292.
- Baron, Reuben M, and David 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), 1173-1182.
- Belia, Sarah, Fiona Fidler, Jennifer Williams, and Geoff Cumming (2005), "Researchers Misunderstand Confidence Intervals and Standard Error Bars," *Psychological Methods*, 10 (4), 389-396.
- Bergkvist, Lars (2015), "Appropriate Use of Single-Item Measures is Here to Stay," *Marketing Letters* (26), 245-255.
- Blanton, Hart and James Jaccard (2006), "Arbitrary Metrics in Psychology," *American Psychologist*, 61 (1), 27-41.
- Bollen, Kenneth A., and Richard Stine (1990), "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability," *Sociological Methodology*, 20, 115-140.

- Bonem, Emily M., Phoebe C. Ellsworth, and Richard Gonzalez (2015), “Age Differences in Risk: Perceptions, Intentions, and Domains,” *Journal of Behavioral Decision Making*, 28 (4), 317-330.
- Bullock, John G., Donald P. Green, and Shang E. Ha (2010), “Yes, But What’s the Mechanism? (Don’t Expect an Easy Answer),” *Journal of Personality and Social Psychology*, 98 (4), 550-558.
- Bullock, John G. and Shang E. Ha (2011), “Mediation Analysis is Harder Than it Looks,” in: James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia (eds.), *Cambridge Handbook of Experimental Political Science*, Cambridge, UK: Cambridge University Press, pp. 508-521.
- Button, Katherine S., John P.A. Ioannides, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S.J. Robinson, and Marcus R. Munafò (2013), “Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience,” *Nature Reviews Neuroscience*, 1-12, doi: 10.1038/nrn3475.
- Cashen, Luke H. and Scott W. Geiger (2004), “Statistical Power and the Testing of Null Hypotheses: A Review of Contemporary Management Research and Recommendations for Future Studies,” *Organizational Research Methods*, 7 (2), 151-167.
- Cheong, JeeWon, David P. MacKinnon, and Siek Toon Khoo (2003), “Investigation of Mediation Processes Using Parallel Process Latent Growth Curve Modeling,” *Structural Equation Modeling*, 10(2), 238-262.
- Cohen, Jacob (1962), “The Statistical Power of Abnormal-Social Psychology Research: A Review,” *Journal of Abnormal and Social Psychology*, 65, 145-153.

- Cohen, Jacob, Patricia Cohen, Stephen G. West, and Leona S. Aiken (2002), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Mahwah, NJ.: Lawrence Erlbaum Associates, 3<sup>rd</sup> edition.
- Cole, David A. and Kristopher J. Preacher (2014), “Manifest Variable Path Analysis: Potentially Serious and Misleading Consequences Due to Uncorrected Measurement Error,” *Psychological Methods*, 19 (2), 300-315.
- Edwards, Jeffrey R. and James W. Berry (2010), “The Presence of Something or the Absence of Nothing: Increasing Theoretical Precision in Management Research,” *Organizational Research Methods*, 13 (4), 668-689.
- Eisend, Martin and Farid Tarrahi (2014), “Meta-Analysis Selection Bias in Marketing Research,” *International Journal of Research in Marketing*, 31, 317-326.
- Fernbach, Philip M., Christina Kan, and John G. Lynch Jr. (2015), “Squeezed: Coping with Constraint through Efficiency and Prioritization,” *Journal of Consumer Research*, 41 (5), 1204-1227.
- Fiedler, Klaus, Malte Schott, and Thorsten Meiser (2011), “What Mediation Analysis Can (Not) Do,” *Journal of Experimental Social Psychology*, 47, 1231-1236.
- Fornell, Claes and David F. Larcker (1981), “Structural Equation Models with Unobserved Variables and Measurement Error: Algebra and Statistics,” *Journal of Marketing Research*, 18, 382-380.
- Fritz, Matthew S. and David P. MacKinnon (2007), “Required Sample Size to Detect the Mediated Effect,” *Psychological Science*, 18 (3), 233-239.
- Gelman, Andrew (2008), “Scaling Regression Inputs by Dividing by Two Standard Deviations,” *Statistics in Medicine*, 27, 2865-2873.

- Gelman, Andrew and John Carlin (2014), "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors," *Perspectives in Psychological Science*, 9(6), 641-651.
- Gelman, Andrew and Hal Stern (2006), "The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant," *American Statistician*, 60 (4), 328-331.
- Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman (2016), "Statistical Tests, *P* values, Confidence Intervals, and Power: A Guide to Misinterpretations," *European Journal of Epidemiology*, 31, 337-350.
- Halquest, Michael (2016), "MplusAutomation: Automatic Mplus Model Estimation and Interpretation," R package version 0.6-4.
- Harris, Christine R., Michael Jenkins, and Dale Glaser (2006), "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?" *Judgment and Decision Making*, 1 (1), 48-63.
- Hayes, Andrew F. (2012), "PROCESS: A Versatile Computational Tool for Observed variable Mediation, Moderation, and Conditional Process Modeling," Retrieved from <http://www.afhayes.com/public/process2012.pdf>.
- Hicks, Raymond and Dustin Tingley (2011), "Causal Mediation Analysis," *The Stata Journal*, 11 (4), 1-15.
- Iacobucci, Dawn, Neela Saldanha, and Xiaoyan Deng (2007), "A Meditation on Mediation: Evidence that Structural Equations Models Perform Better than Regressions," *Journal of Consumer Psychology*, 17 (2), 140-154.
- Imai, Kosuke, Luke Keele, and Dustin Tingley (2010), "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 15 (4), 309-334.

- Imai, Kosuke, Luke Keele, and Teppei Yamamoto (2010), "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25 (1), 51-71.
- Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto (2013), "Experimental Designs for Identifying Causal Mechanism," *Journal of the Royal Statistical Society A*, 176 (1), 5-51, with discussion.
- Ioannidis, John P.A. (2005), "Why Most Published Research Findings are False," *PLOS Medicine*, 2 (8), 696-701.
- Jhang, Ji Hoon and John G. Lynch Jr. (2015), "Pardon the Interruption: Goal Proximity, Perceived Spare Time, and Impatience," *Journal of Consumer Research*, 41 (5), 1267-1283.
- Judd, Charles M. and David Kenny (1981), "Process Analysis: Estimating Mediation in Treatment Evaluations," *Evaluation Review*, 5, 602-619.
- Kenny, David A. and Charles M. Judd (2014), "Power Anomalies in Testing Mediation" *Psychological Science*, 25(2), 334-339.
- Kerr, Norbert L. (1998), "HARKing: Hypothesizing After the Results Are Known," *Personality and Social Psychology Review*, 2, 196-217.
- Kim, Jae-On and G. Donald Ferree, Jr. (1981), "Standardization in Causal Analysis," *Sociological Methods & Research*, 10 (2), 187-210.
- Kim, Sara and Ann L. McGill (2011), "Gaming with Mr. Slot or Gaming the Slot Machine? Power, Anthropomorphism, and Risk Perception," *Journal of Consumer Research*, 38 (1), 94-107.
- Kirca, Ahmet H., Satish Jayachandran, and William O. Bearden (2005), "Market Orientation: A Meta-Analytic Review and Assessment of Its Antecedents and Impact of Performance," *Journal of Marketing*, 69 (April), 24-41.

- Kisbu-Sakarya, Yasemin, David P. MacKinnon, and Milica Miočević (2014), "The Distribution of the Product Explains Normal Theory Mediation Confidence Interval Estimation," *Multivariate Behavioral Research*, 49 (3), 261-268.
- Kupor, Daniella M. and Zakary L. Tormala (2015), "Persuasion, Interrupted: The Effect of Momentary Interruptions on Message Processing and Persuasion," *Journal of Consumer Research*, 42 (2), 300-315.
- Ledgerwood, Alison and Patrick E. Shrout (2011), "The Trade-Off Between Accuracy and Precision in Latent Variable Models of Mediation Processes," *Journal of Personality and Social Psychology*, 101 (6), 1174-1188.
- Lerner, Jennifer S., Ye Li, Piercarlo Valdesolo, and Karim S. Kassam (2015), "Emotion and Decision Making," *Annual Review of Psychology*, 66, 799-823.
- Lisjak, Monika, Andrea Bonezzi, Soo Kim, and Derek D. Rucker (2015), "Perils of Compensatory Consumption: Within-Domain Compensation Undermines Subsequent Regulation," *Journal of Consumer Research*, 41 (5), 1186-1203.
- Lykken, David T. (1991), What's Wrong with Psychology Anyway? In: Dante Cicchetti and William M. Grove (Eds.), *Thinking Clearly about Psychology, Volume 1: Matters of Public Interest*, Minneapolis: University of Minnesota Press, pp. 3-39.
- MacCallum, Ronald C. and James T. Austin (2000), "Applications of Structural Equations Modeling in Psychological Research," *Annual Review of Psychology*, 51, 201-236.
- MacKinnon, David P. (2008), *Introduction to Statistical Mediation Analysis*, New York: Lawrence Erlbaum Associates.

- MacKinnon, David P., JeeWon Cheong, and Angela G. Pirlott (2012), Statistical Mediation Analysis, in: Harris Cooper (ed.), *APA Handbook of Research Methods in Psychology, Vol. 2, Research Designs*, New York: American Psychological Association, 313-331.
- Marszalek, Jacob M., Carolyn Barber, Julie Kohlhart, and Cooper B. Holmes (2011), "Sample Size in Psychological Research over the Past 30 Years," *Perceptual and Motor Skills*, 112 (2), 331-348.
- Mauro, Robert (1990), "Understanding L.O.V.E. (Left Out Variables Error): A Method for Estimating the Effects of Omitted Variables," *Psychological Bulletin*, 108 (2), 314-329.
- Maxwell, Scott E. (2004), "The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies," *Psychological Methods*, 9 (2), 147-163.
- Meehl, Paul E. (1967), "Theory-Testing in Psychology and Physics: A Methodological Paradox," *Philosophy of Science*, 34 (2), 103-115.
- Meehl, Paul E. (1990), "Why Summaries of Research on Psychological Theories are often Uninterpretable," *Psychological Reports*, 66, 195-244.
- Miočević, Milica, Holly P. O'Rourke, David P. MacKinnon, and Henricks C. Brown (2017), "Statistical Properties of Four Effect-Size Measures for Mediation Models," *Behavior Research Methods*, DOI 10.3758/s13428-017-0870-1, published online: 24 march 2017.
- Muthén, Bengt and Tihomir Asparouhov (2015), "Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables," *Structural Equation Modeling*, 22, 12-23.
- Muthén, Linda and Bengt O. Muthén (2015), *MPlus User's Guide*, Seventh Edition. Los Angeles, CA: Muthén and Muthén.



- Muthén, Bengt, Linda Muthén, and Tihomor Asparouhov (2016), *Regression and Mediation Analysis using Mplus*, Los Angeles, CA: Muthén and Muthén.
- Nieuwenhuis, Sander, Birte U. Forstmann, and Eric-Jan Wagenmakers (2011), “Erroneous Analyses of Interactions in Neuroscience: A Problem of Significance,” *Nature Neuroscience*, 14 (9), 1105-1107.
- Open Science Collaboration (2015), “Estimating the Reproducibility of Psychological Science,” *Science*, 349, 1-8, doi: 10.1126/science.aac4716.
- O’Rourke, Holly P. and David P. MacKinnon (2015), “When the Test of Mediation is More Powerful than the Test of the Total Effect,” *Behavioral Research Methods*, 47, 424-442.
- Paolacci and Jesse Chandler (2014), “Inside the Turk: Understanding Mechanical Turk as a Participant Pool,” *Current Directions in Psychological Science*, 23 (3), 184-188.
- Pearl, Judea (2009), “Causal Inference in Statistics: An Overview,” *Statistics Surveys*, 3, 96-146.
- Peterson, Robert A. (1994), “A Meta-Analysis of Cronbach’s Coefficient Alpha,” *Journal of Consumer Research*, 21 (2), 381-391.
- Petrocelli, John V., Joshua J. Clarkson, Melanie B. Whitmire, and Paul E. Moon (2013), “When  $ab \neq c-c'$ : Published Errors in the Reports of Single-Mediator Models,” *Behavior Research Methods*, 45, 595-601.
- Pieters, Rik (2017), *Mediation Analysis: Inferring Causal Processes in Marketing From Experiments*, Peter S.H. Leeflang et al. (eds.), *Advanced Methods in Modeling Markets*, Springer International, Chapter 8, forthcoming.
- Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff (2012), “Sources of Method Bias in Social Science Research and Recommendations on How to Control it,” *Annual Review of Psychology*, 63, 539-569.

- Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff (2016), "Recommendations for Creating Better Concept Definitions in the Organizational, Behavioral, and Social Sciences," *Organization Research Methods*, 19(2), 159-203.
- Preacher, Kristopher J. (2015), "Advances in Mediation Analysis: A Survey and Synthesis of New Developments," *Annual Review of Psychology*, 66, 825-852.
- Preacher, Kristopher J. and Andrew F. Hayes (2008), "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models," *Behavior Research Methods*, 40 (3), 879-891.
- Preacher, Kristopher and Ken Kelley (2011), "Effect Size Measures for Mediation Models: Quantitative Strategies for Communicating Indirect Effects," *Psychological Methods*, 16 (2), 93-115.
- Preacher, Kristopher, Derek D. Rucker, and Andrew Hayes (2007), "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions," *Multivariate Behavioral Research*, 42 (1), 185-227.
- Richard, F.D., Charles F. Bond, Jr., and Juli Stoles-Zoota (2003), "One Hundred Years of Social Psychology Quantitatively Described," *Review of General Psychology*, 7 (4), 331-363.
- Richardson, Hettie A., Marcia J. Simmering, and Michael C. Sturman (2009), "A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance," *Organizational Research Methods*, 12, 762-800.
- Roberts, Seth and Harold Pashler (2000), "How Persuasive is a Good Fit? A Comment on Theory Testing," *Psychological Review*, 107 (2), 358-367.

- Rucker, Derek D., Kristopher J. Preacher, Zakary L. Tormala, and Richard E. Petty (2011), "Mediation Analysis in Social Psychology: Current Practices and New Recommendations," *Social and Personality Psychology Compass*, 5/6, 359-371.
- Sawyer, Alan G. and A. Dwayne Ball (1981), "Statistical Power and Effect Size in Marketing Research," *Journal of Marketing Research*, 18, 275-290.
- Scheker, Nathaniel and Jane F. Gentleman (2001), "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals," *The American Statistician*, 55 (3), 182-186.
- Schiff, Rom Y. and Moty Amar (2015), "Pain and Preferences: Observed Decisional Conflict and the Convergence of Preferences," *Journal of Consumer Research*, 42 (4), 515-534.
- Sedlmeier, Peter and Gerd Gigerenzer (1989), "Do Studies of Statistical Power have an Effect on the Power of Studies?," *Psychological Bulletin*, 105 (2), 309-316.
- Sha, Avni M., Noah Eisenkraft, James R. Bettman, and Tanya L. Chartrand (2016), "'Paper or Plastic?' How We Pay Influences Post-Transaction Connection," *Journal of Consumer Research*, 42 (5), 688-708.
- Shook, Christopher, L., David J. Ketchen, J.R., G. Tomas Hult, and K. Michele Kacmar (2004), "An Assessment of the Use of Structural Equation Modeling in Strategic Management Research," *Strategic Management Journal*, 25, 397-404.
- Shrout, Patrick E. and Niall Bolger (2002), "Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations," *Psychological Methods*, 7 (4), 422-445.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011), "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 22, 1359-1366.

- Smith, Eliot R. (1982), "Beliefs, Attributions, and Evaluations: Nonhierarchical Models of Mediation in Social Cognition," *Journal of Personality and Social Psychology*, 43 (2), 248-259.
- Smith, Rosanna K., George E. Newman, and Ravi Dhar (2016), "Closer to the Creator: Temporal Contagion Explains the Preference for Earlier Serial Numbers," *Journal of Consumer Research*, 42 (5), 653-668.
- Soster, Robin L., Andrew D. Gershoff, and William O. Bearden (2014), "The Bottom Dollar Effect: The Influence of Spending to Zero on Pain of Payment and Satisfaction," *Journal of Consumer Research*, 41(3), 656-677.
- Spearman, C. (1904), "The Proof and Measurement of Association between Two Things," *The American Journal of Psychology*, 15 (1), 72-101.
- Spencer, Steven J., Mark P. Zanna, and Geoffrey T. Fong (2005), "Establishing a Causal Chain: Why Experiments are Often More Effective than Mediational Analyses in Examining Psychological Processes," *Journal of Personality and Social Psychology*, 89 (6), 845-851.
- Ten Have, Thomas R. and Marshall M. Joffe (2010), "A Review of Causal Estimation of Effects in Mediation Analyses," *Statistical Methods in Medical Research*, 21, 77-107.
- Thoemmes, Felix (2015), "Reversing Arrows in Mediation Models Does Not Distinguish Plausible Models," *Basic and Applied Social Psychology*, 37, 226-234.
- Valeri, Linda and Tyler VanderWeele (2013), "Mediation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros," *Psychological Methods*, 18(2), 137-150.

- Valsesia, Francesca, Joseph C. Nunes, and Andrea Ordanini (2016), "What Wins Awards is Not Always What I Buy: How Creative Control Affects Authenticity and Thus Recognition (But Not Liking)," *Journal of Consumer Research*, 42 (6), 897-914.
- VanderWeele, Tyler J., Linda Valeri, and Elizabeth L. Ogburn (2012), "The Role of Measurement Error and Misclassification in Mediation Analysis," *Epidemiology*, 23 (4), 561-564.
- Voorhees, Clay M., Michael K. Brady, Roger Calantone, and Edward Ramirez (2016), "Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies," *Journal of the Academy of Marketing Science*, 44, 119-134.
- Wanous, John. P. and Michael J. Hudy (2001), "Single-Item Reliability: A Replication and Extension," *Organizational Research Methods*, 4 (4), 361-375.
- Westfall, Jacob, and Tal Yarkoni (2016), "Statistically Controlling for Confounding Constructs is Harder than You Think," *PLOS ONE*, March 2016, 1-22.
- Williams, Larry J. (2012), Equivalent Models: Concepts, Problems, and Alternatives, in Rick H. Hoyle (ed.), *Handbook of Structural Equation Modeling*, New York, The Guilford Press, pp. 247-260.
- Wright, Sewall (1921), "Correlation and Causation," *Journal of Agricultural Research*, 20 (7), 557-585.
- Yuan, Ying and David P. MacKinnon (2009), "Bayesian Mediation Analysis," *Psychological Methods*, 14 (4), 301-322,
- Zhao, Xinshu, John Lynch, Jr., and Qimei Chen (2010), "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis," *Journal of Consumer Research*, 37 (2), 197-206.

Zhang, Jie, Michel Wedel, and Rik Pieters (2009), "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis," *Journal of Marketing Research*, 46 (3), 669-681.

**TABLE 1**  
**MEDIATION ANALYSIS IN *JOURNAL OF CONSUMER RESEARCH* VOL. 41 AND 42**

Category	Results
Number of articles	138
Number of articles with experiment	121 (88%)
Number of articles with experiment and mediation	86 (71%)
Number of mediation analyses	166 ( $M = 1.93$ , range 1 - 7)
<i>Mediation Model (n = 166):</i>	
Basic mediation	33%
Multiple mediation	17%
Moderated mediation	49%
<i>Samples and Measures:</i>	
Sample from MTurk	73 (44%)
Sample size in mediation analysis	$M = 183$ ( $SD = 124$ , range 35 - 1214)
<i>Mediator and Outcome Measures:</i>	
Number of items for mediator ( $n = 151$ )	$M = 2.94$ ( $SD = 3.09$ , range 1 - 25)
Single item for mediator	43%
Reliability (alpha) of multi-item mediator (for $n = 69$ )	$M = .84$ ( $SD = .11$ , range .51 - .97)
Number of items for outcome ( $n = 141$ )	$M = 2.28$ ( $SD = 3.39$ , range 1 - 30)
Single item for outcome	64%
Reliability (alpha) of multi-item outcome (for $n = 45$ )	$M = .85$ ( $SD = .10$ , range .53 - .99)
Single item for mediator and outcome ( $n = 143$ )	34%
<i>Conditions Considered (n = 166):</i>	
Directionality: Equivalent models acknowledged	8
Reliability: Measurement error accounted	0
Unconfoundedness: Omitted variables examined	2
Distinctiveness: Discriminant validity established	3
Power: Statistical power of testing reported	0
<i>Communication of Results (n = 166):</i>	
Indirect effect (a x b): Estimate, and Confidence Interval	61%, and 94%
Conditional direct effect (c')	37%
Indirect effect (a x b) and conditional direct effect (c')	25%
b-path (between M and Y)	34%
Test of moderated mediation ( $a3 \times b$ ) (for $n = 82$ )	63%

NOTE.—  $n$  items and reliability of mediator and outcome could not always be determined from study descriptions. Analyses with combination of multiple and moderated mediation were coded as moderated mediation.

**TABLE 2**  
**MEANINGFUL MEDIATION ANALYSIS**

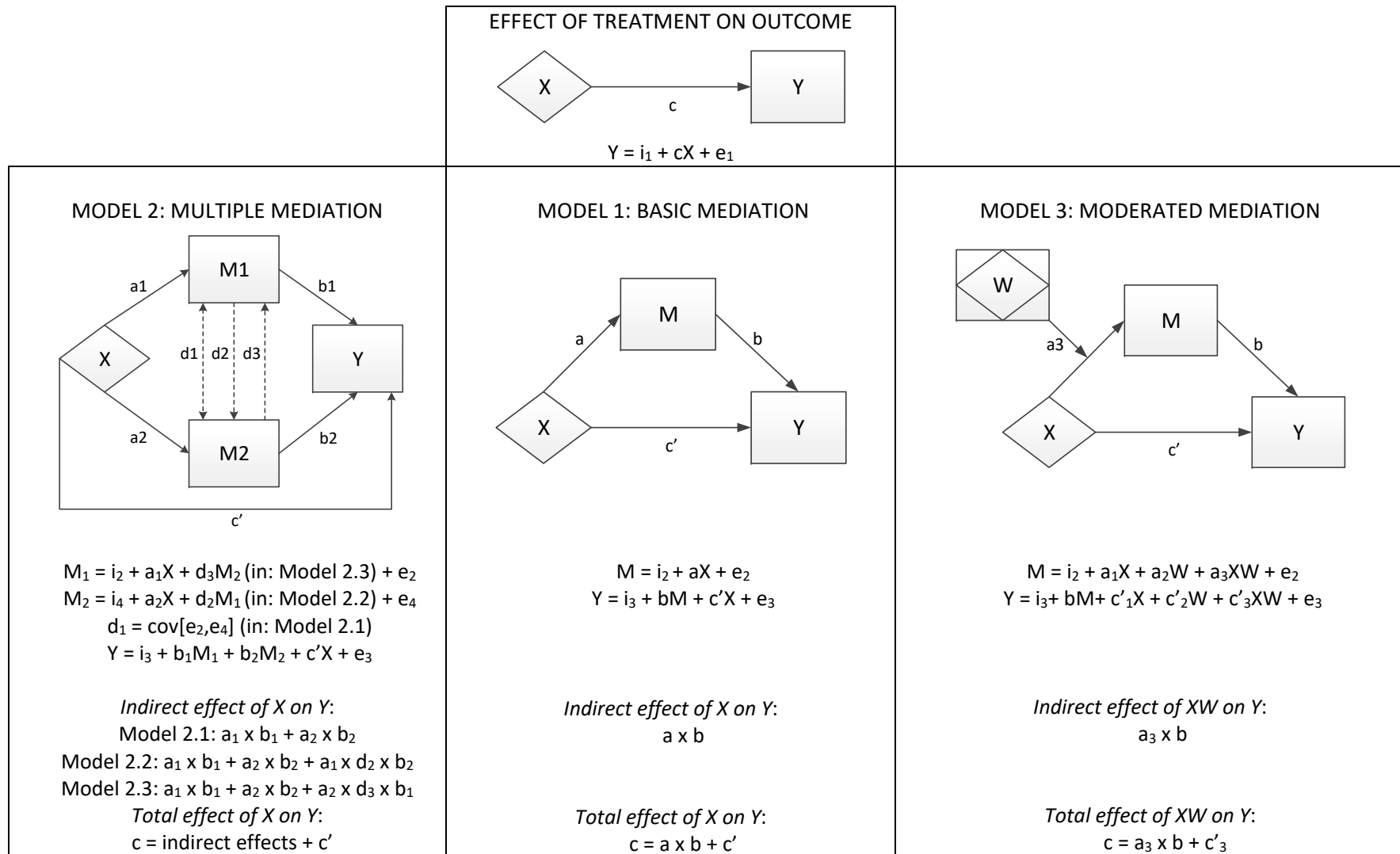
Condition	Description	Recommendation
1. Directionality	Most plausible causal direction is from mediator to outcome, and as specified between mediators	<ul style="list-style-type: none"> <li>a. Provide strong evidence from logic, theory, and prior research that the hypothesized causal direction is more plausible than indicated alternatives.</li> <li>b. If 1a fails: Refrain from statistical mediation analysis, and consider Ms and Y as separate, correlated effects of X.</li> <li>c. Do longitudinal mediation analysis by measuring Ms and Ys, and potentially manipulating X, repeatedly over time.</li> <li>d. Do experimental mediation analysis by (also) manipulating Ms.</li> </ul>
2. Reliability	Measurement error in mediator(s) and outcome is ignorable	<ul style="list-style-type: none"> <li>a. Improve reliability of measures of Ms and Y.</li> <li>b. Account for unreliability of measures of Ms and Y, by using a structural equation model.</li> </ul>
3. Unconfoundedness	Effect of unobserved variables on the association between mediator (s) and outcome is ignorable	<ul style="list-style-type: none"> <li>a. Reduce common-method bias by using diverse methods to measure Ms and Y, and by increasing their temporal and spatial distance.</li> <li>b. Account for potential theory-based confounders of correlation between Ms and Y.</li> <li>c. Do a sensitivity analysis of potential omitted variable bias on the correlation between Ms and Y.</li> <li>d. Do 1c or 1d.</li> </ul>
4. Distinctiveness	Mediator(s) and outcome are theoretically and empirically distinct variables	<ul style="list-style-type: none"> <li>a. Examine conceptually distant Ms and Ys</li> <li>b. Do 2a, 3a, and 3b.</li> <li>c. Increase the sample size.</li> <li>d. Provide evidence for discriminant validity of Ms and Y.</li> <li>e. If 4d fails: Refrain from statistical mediation analysis, and consider Ms and Y as indicators of a single effect of X.</li> </ul>
5. Power	Statistical power is sufficient to identify true non-null direct and indirect effects	<ul style="list-style-type: none"> <li>a. Do 2a. and 2b.</li> <li>b. Do 4c.</li> <li>c. Provide evidence of sufficient statistical power to identify hypothesized effects.</li> </ul>
6. Mediation	The treatment has an indirect effect on the outcome via the mediator(s)	Draw conclusions about size, sign, and significance of indirect and conditional direct treatment effects, <i>after</i> the first five conditions are met.



**TABLE 3**  
**MEANINGFUL COMMUNICATION OF MEDIATION ANALYSIS**

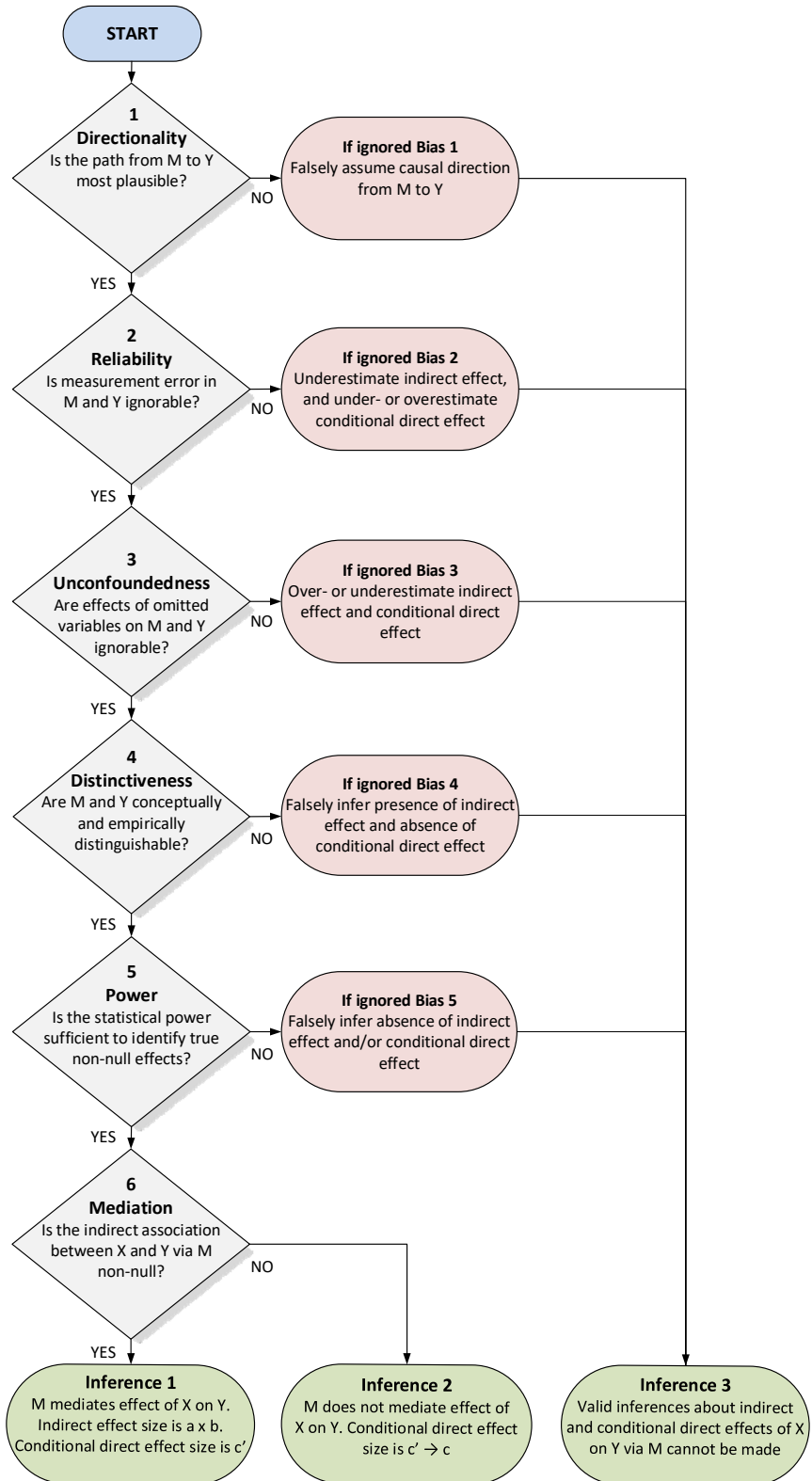
Component	Description	Recommendation
1. Effect Decomposition	Mediation analysis decomposes the total treatment effect into indirect and conditional direct effects	Report point estimates and their uncertainty (standard error, 95% CI and/or <i>p</i> -value) for <i>all</i> paths, for the indirect, conditional direct, and total treatment effect, preferably in one or a few tables and/or in boxes-and-arrows format.
2. Effect Size	Standardized effect sizes indicate meaningfulness of effects, and facilitate within- and between study comparisons	<ul style="list-style-type: none"> <li>a. Report standardized effect sizes (e.g., by standardizing all variables except X).</li> <li>b. Report unstandardized effect sizes if mediator and outcome measures are on non-arbitrary, meaningful response scales.</li> <li>c. Consider reporting the percentage mediated.</li> <li>d. Declare what is being reported (standardized, unstandardized, both).</li> </ul>
3. Difference Testing	Tests of differences between hypothesized effects support accurate inferences about the differences	<p><i>For basic mediation:</i></p> <ul style="list-style-type: none"> <li>a. If a difference between indirect and conditional direct effect is hypothesized, report a test of it.</li> </ul> <p><i>For multiple mediation:</i></p> <ul style="list-style-type: none"> <li>b. If differences between indirect effects are hypothesized, report tests of these.</li> </ul> <p><i>For moderated mediation:</i></p> <ul style="list-style-type: none"> <li>c. Report tests of the hypothesized moderated mediation effect and/or report tests of differences between indirect effects for levels of the moderator.</li> </ul>
4. Data Sharing	Raw and summary statistics data enable secondary and meta-analysis, and knowledge accumulation more generally	<ul style="list-style-type: none"> <li>a. Report summary statistics data (SSD) in the manuscript: Sample size, means, SD, correlations, reliabilities of multi-item measures, and coding of <i>all</i> variables.</li> <li>b. Provide access to the raw data.</li> </ul>

**FIGURE 1**  
**THREE MEDIATION MODELS**

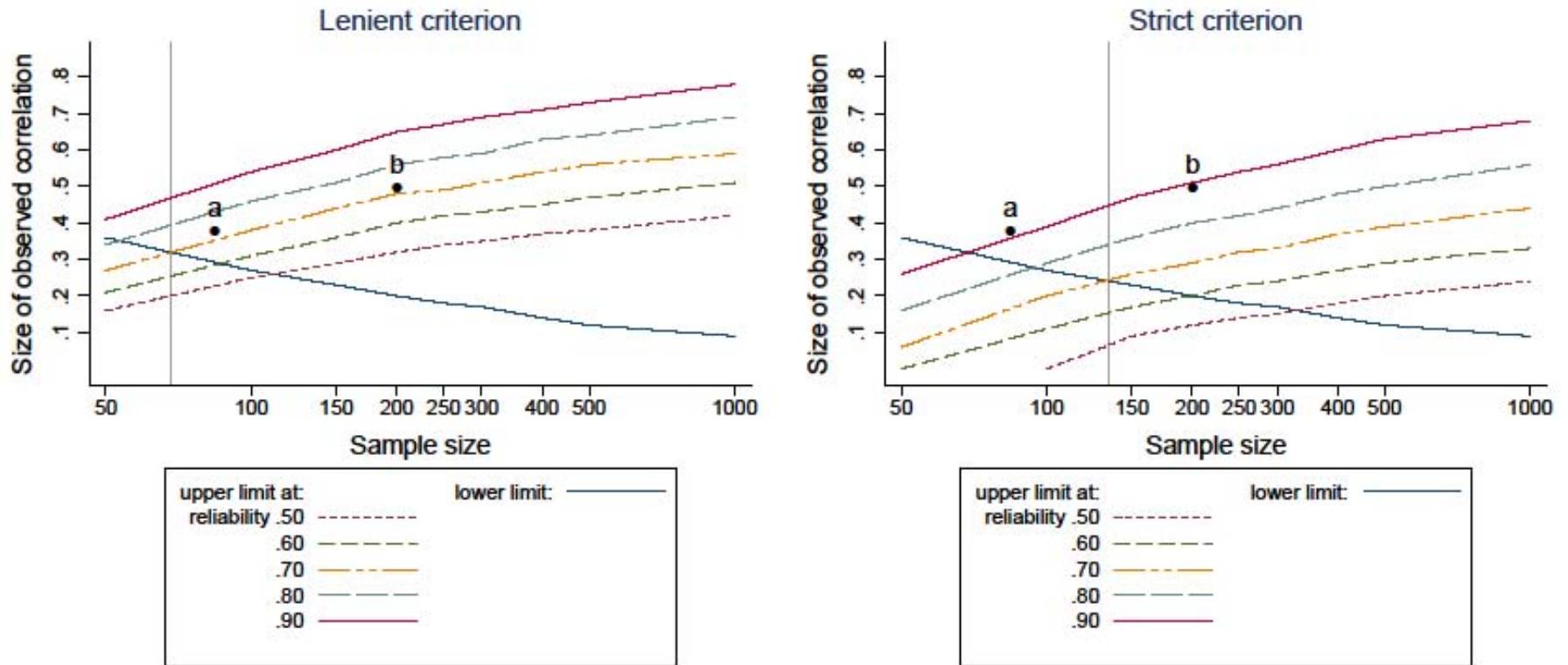


NOTE.—Diamond denotes treatment with random assignment to experimental conditions (X). Boxes are observed measures of Mediator (M) and Outcome (Y), assumed to be continuous. Moderators (W) are experimental conditions (diamond) or measured differences between participants (square). Single-headed arrows indicate causal direction. Double-headed arrow is a covariance.  $\text{cov}[e_2, e_4]$  is the covariance between residuals of M1 and M2. Each d-link (1 to 3) is included in only one multiple mediation model: d1 in Model 2.1, d2 in Model 2.2, and d3 in Model 2.3.

**FIGURE 2**  
SIX CONDITIONS FOR MEANINGFUL MEDIATION ANALYSIS



**FIGURE 3**  
THE SWEETSPOT: REGION OF STATISTICALLY MEANINGFUL MEDIATOR-OUTCOME CORRELATIONS



NOTE.—Horizontal axis is sample size, on a log-scale. Vertical axis is observed mediator-outcome correlation, when not accounting for measurement error. Lower curve is minimum correlation significantly different from zero ( $p < .05$ ) at 80% power. Upper curves are maximum observed correlations meeting the respective discriminant validity criterion, for a specific reliability. Lenient criterion (left plot) is met when the mediator-outcome correlation is less than one. Strict criterion (right plot) is met when smallest reliability of mediator and outcome is larger than squared true mediator-outcome correlation. Reliability for lenient criterion is geometric mean of the two reliabilities ( $\sqrt{(r_{MM}r_{YY})}$ ). Reliability for strict criterion is size of the smallest reliability. The Sweetspot for reliabilities of .70 starts to the right of the vertical line. Bullet (a) is correlation .38, sample size 84. Bullet (b) is correlation .50, sample size 200.

## HEADINGS LIST

- 1) INTRODUCTION
- 1) MEDIATION ANALYSIS
- 1) MEANINGFUL MEDIATION ANALYSIS
  - 2) The Directionality Condition
  - 2) The Reliability Condition
  - 2) The Unfoundedness Condition
  - 2) The Distinctiveness Condition
  - 2) The Power Condition
  - 2) The Mediation Condition
- 1) THE SWEETSPOT
- 1) MEANINGFUL COMMUNICATION OF MEDIATION
  - 2) Effect Decomposition
  - 2) Effect Size
  - 2) Difference Tasting
  - 2) Data Sharing
- 1) CONCLUSION
  - 2) Conditions and Communication
  - 2) Future Work
- 1) APPENDIX A
- 1) APPENDIX B
- 1) REFERENCES