JCP Journal of Consumer Psychology

SCP Society for Consumer Psychology

Meta-analysis of Studies with Multiple Contrasts and Differences in Measurement Scales

Blakeley B. McShane, and Ulf Böckenholt
Northwestern University

Accepted by Anirban Mukhopadhyay, Editor; Associate Editor, Joel Huber

The common approach to meta-analysis is overwhelmingly dominant in practice but suffers from a major limitation: It is suitable for analyzing only a single effect of interest. However, contemporary psychological research studies—and thus meta-analyses of them—typically feature multiple dependent effects of interest. In this paper, we introduce novel meta-analytic methodology that (a) accommodates an arbitrary number of effects—specifically, contrasts of means—and (b) yields results in standard deviation units in order to adjust for differences in the measurement scales used for the dependent measure across studies. Importantly, when all studies follow the same two-condition study design and interest centers on the simple contrast between the two conditions as measured on the standardized mean difference (or Cohen's d) scale, our approach is equivalent to the common approach. Consequently, our approach generalizes the common approach to accommodate an arbitrary number of contrasts. As we illustrate and elaborate on across three extensive case studies, our approach has several advantages relative to the common approach. To facilitate the use of our approach, we provide a website that implements it.

Keywords standardized mean difference; Cohen's d; heterogeneity; between-study variation; hierarchical; meta-analysis; multilevel; random effects

Introduction

The common approach to meta-analysis involves collapsing the individual-level observations from each study into a single effect of interest; converting these effects to a standardized scale such as the standardized mean difference (or Cohen's d) scale or the correlation scale in order to adjust for differences in the measurement scales used for the dependent measure across studies; and fitting the basic random effects meta-analytic model to the converted effects. Although this approach is overwhelmingly dominant in practice, it suffers from a major limitation: It is suitable for analyzing only a single effect of interest.

However, contemporary psychological research studies—and thus meta-analyses of them—typically feature multiple dependent effects of interest. To accommodate this, standard practice involves applying the common approach separately to each effect in turn. However, it would be preferable to model all effects jointly in a single analysis.

To facilitate this, we introduce novel meta-analytic methodology that (a) accommodates an arbitrary number of effects—specifically, contrasts of means (e.g., of the multiple study conditions that arise from the variation of one or more experimental factors such as simple effects, main effects, and interaction effects)—and (b) yields results in standard deviation units in order to adjust for differences in the measurement scales used for the dependent measure across studies. Importantly, when all studies follow the same two-condition study design and interest centers on the simple contrast between the two conditions as measured on the standardized mean difference scale, our approach—which we hereafter refer to as the Multiple Contrast Standardized Meta-analysis (MCSM) approach—is equivalent to the common approach—which we hereafter refer to as the Single Contrast Standardized Meta-analysis (SCSM) approach. Consequently, the MCSM approach generalizes the SCSM approach to accommodate an arbitrary number of contrasts.

As we illustrate and elaborate on across three extensive case studies, the MCSM approach has

Received 30 June 2018; accepted 28 February 2021
Available online 9 March 2021

*Correspondence concerning this article should be addressed to Blakeley B. McShane, Marketing Department, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208. Electronic mail may be sent to b-mcshane@kellogg.northwestern.edu.

 1057-7408/2021/1532-7663/32(1)/23–40
DOI: 10.1002/jcpy.1236

several advantages relative to the SCSM approach. First, the MCSM approach is user-friendly because it requires only basic summary information (e.g., means, standard deviations, and sample sizes) from each study condition. In contrast, the SCSM approach requires preprocessing of the data to convert effects to a standardized scale.

Additional advantages follow directly from the fact that the MCSM approach models the data from all conditions of all studies jointly in a single analysis whereas the SCSM approach models various subsets of the data in turn across multiple analyses. Specifically, the MCSM approach yields more accurate estimates of heterogeneity (e.g., between-study variation in effects) as compared to the SCSM approach. This in turn results in more accurate estimates of the standard errors of the estimates of the effects and therefore improved statistical inference (e.g., confidence intervals, *p*-values). Further, the MCSM approach yields estimates of the covariance matrix of the estimates of the effects while the SCSM approach cannot. Finally, the MCSM approach yields coherent estimates of the effects while the SCSM approach does not.

A final advantage is that the MCSM approach accommodates (a) a mix of study designs (e.g., two-condition and two-by-two study designs; between-subjects and within-subjects study designs) and (b) study-level moderators (or covariates).

In the remainder of this paper, we first describe the MCSM approach. We then illustrate it in the context of three case studies. We conclude by discussing some cautions relevant when applying the MCSM approach and providing a brief summation.

To facilitate the use of the MCSM approach, we provide a website that implements it: https://blakemcshane.shinyapps.io/mcsmeta/. The website includes a detailed tutorial that shows how to replicate the case studies presented in this paper and how to apply it to other studies.

Model Description

The MCSM approach generalizes the SCSM approach to accommodate an arbitrary number of contrasts. It also accommodates a mix of study designs and study-level moderators.

To achieve this, the MCSM approach decomposes each observation—that is, the mean of the individual-level observations in each condition of each study—into four components: (a) an overall study average component, (b) an overall condition average component, (c) a study condition component that reflects heterogeneity, and (d) a study condition component that reflects sampling error. The first component reflects the measurement scales used for the dependent measure in and other aspects specific to each study. The second component reflects the experimental factors specific to each condition. The third component reflects heterogeneity specific to each study condition (and, when one or more studies follow a within-subjects study design, to each subject group). The MCSM approach assumes heterogeneity and sampling error operate independently; it further assumes they have zero mean as the mean is captured by the overall study and overall condition average components.

To accommodate study-level moderators, the MCSM approach introduces additional components to reflect the association of each moderator with each condition.

Importantly, despite requiring only basic summary information, the model underlying the MCSM approach is equivalent to that underlying the "gold standard" meta-analytic approach—namely, an appropriately specified hierarchical (or multilevel) model fit to the individual-level observations (Cooper & Patall, 2009; Haidich, 2010; McShane & Böckenholt, 2020; Simmonds et al., 2005; Stewart & Tierney, 2002).

We refer the reader interested in the full details regarding the MCSM model specification and estimation procedure to the Appendix.

Case Studies

In this section, we illustrate the MCSM approach in the context of three case studies. The first, based on unconscious thought theory, illustrates the advantages that the MCSM has relative to the SCSM approach by directly comparing the two approaches. The second, based on the choice overload hypothesis, more deeply illustrates how the MCSM approach accommodates a mix of study designs by further introducing a mix of between-subjects and within-subjects study designs. Finally, the third, based on the flexible correction model, illustrates how the MCSM approach accommodates study-level moderators.

In all three cases studies, we use hypothetical data. This hypothetical data—which appears in Table 1, Table 3, and Table 5—is based off the results presented in Bos, Dijksterhuis, and van Baaren (2011), McShane and Böckenholt (2018), and Petty, Wegener, and White (1998), respectively. Code to generate this hypothetical data is available in our Methodological Details Appendix.

Table 1
Unconscious Thought Theory Study Summary Information

Study	Condition	Mean	Standard deviation	Sample size
Study 1	Immediate	7.1600	6.8254	50
	Unconscious	11.3400	7.4741	50
	Distraction			
Study 2	Immediate	3.8600	3.2225	100
	Unconscious	5.8300	3.3456	100
	Distraction			
Study 3	Immediate	45.6250	40.0715	40
	Unconscious	60.4250	39.6303	40
	Distraction			
Study 4	Immediate	4.0267	3.3287	75
	Unconscious	5.5867	3.4134	75
	Distraction	5.1867	3.3559	75
Study 5	Immediate	7.1400	6.4460	50
	Unconscious	14.3800	6.1740	50
	Distraction	7.8400	7.3049	50
Study 6	Immediate	4.2700	3.1777	100
	Unconscious	5.4700	3.5745	100
	Distraction	5.0800	3.2370	100

Case Study I: Unconscious Thought Theory

A researcher was interested in conducting a meta-analysis of studies of the conjecture that unconscious thought leads to an automatic weighting process as compared to immediate decision-making such that important decision attributes receive more weight and unimportant decision attributes receive less weight. Toward this end, the researcher surveyed the literature and identified as relevant six hypothetical studies.

Three of the studies followed a two-condition study design (immediate decision or unconscious thought) while the other three followed a three-condition study design (immediate decision or unconscious thought or mere distraction) intended to refute an alternative account. Further, all of the studies followed a between-subjects study design. Finally, the dependent measure in these studies was a product quality rating composite which was measured on a twenty-point integer scale in Study 1 and Study 5; a ten-point integer scale in Study 2, Study 4, and Study 6; and a one hundred point integer scale in Study 3.

The effects of interest were (a) the simple effect of unconscious thought versus immediate decision, (b) the simple effect of unconscious thought versus mere distraction, and (c) the simple effect of mere distraction versus immediate decision, which are given by the respective contrast vectors (−1 1 0), (0 1 −1), and (−1 0 1). The first two effects were predicted to be positive (i.e., because higher composite scores indicate more appropriate weighting) while the third effect was predicted to be null.

Summary information for these studies (specifically the mean, standard deviation, and sample size of the individual-level observations in each condition of each study) can be found in Table 1. We note that here and hereafter we provide excess digits to facilitate the ability of the reader to reproduce our results via our website.

We begin by discussing the SCSM approach. We next discuss the MCSM approach. Finally, we discuss the estimates obtained via the two approaches and review how this case study illustrates several advantages that the MCSM approach has relative to the SCSM approach.

Given an effect of interest, the SCSM approach involves five steps. First, one computes the pooled standard deviation of each study. Second, one computes the contrast corresponding to the effect for each study thereby collapsing the individual-level observations from each study into a single effect. Third, one divides the contrast for each study by the pooled standard deviation of the study thereby converting these effects to a standardized scale in order to adjust for differences in the measurement scales used for the dependent measure across studies. Fourth, one computes the sampling standard deviation of each converted effect for each study. Fifth, one fits the basic random effects meta-analytic model to the converted effects. When there is more than one effect of interest, the second through fifth steps are repeated separately for each effect in turn.

We illustrate these steps using the basic summary information found in Table 1 and the simple effect of unconscious thought versus immediate decision as the effect of interest. First, one computes the pooled standard deviation of each study; for example, for Study 1, this is given by

$$\sqrt{\frac{\sum_i (n_i-1)\sigma_i^2}{\sum_i (n_i-1)}} = \sqrt{\frac{(50-1)\cdot 6.8254^2 + (50-1)\cdot 7.4741^2}{(50-1)+(50-1)}} = 7.1571$$

where σ_i and n_i denote the respective standard deviation and sample size of the individual-level observations in condition i of the study. Second, one computes the contrast corresponding to the simple effect of unconscious thought versus immediate decision for each study; for example, for Study 1, this is given by $(11.3400 - 7.1600) = 4.1800$.

Third, one divides the contrast for each study by the pooled standard deviation of the study; for example, for Study 1, this is given by $4.1800/7.1571 = 0.5840$. Fourth, one computes the sampling standard deviation of each converted effect for each study; for example, for Study 1, this is given by $\sqrt{1/50 + 1/50} = 0.2000$. One repeats these four steps for each study and then fits the basic random effects meta-analytic model to the six converted effects (i.e., one from each study). One then repeats the second through fifth steps for the simple effect of unconscious thought versus mere distraction and then again for the simple effect of mere distraction versus immediate decision, noting that the basic random effects meta-analytic model for these two effects is fit to only three converted effects (i.e., one each from Study 4, Study 5, and Study 6) due to the study designs.

In contrast, the MCSM approach involves simply fitting the model detailed in the Appendix directly to basic summary information from each study condition like that found in Table 1 regardless of the identity or number of effects of interest.

Results from the two approaches can be found in Table 2. We begin by discussing those from the SCSM approach. As can be seen, the point estimate of the simple effect of unconscious thought versus immediate decision is 0.5662 (in standard deviation units), with a 95% confidence interval estimate of [0.3684, 0.7641] (i.e., $0.5662 \pm 1.9600 \cdot 0.1010$); the point estimate of the simple effect of unconscious thought versus mere distraction is 0.3928, with a 95% confidence interval estimate of [−0.1585, 0.9440]; and the point estimate of the simple effect of mere distraction versus immediate decision is 0.2462, with a 95% confidence interval estimate of [0.0614, 0.4310]. These estimates are not fully consistent with the predictions of unconscious thought theory.

We now discuss assessments of heterogeneity from the SCSM approach. There are three standard assessments of heterogeneity associated with the SCSM approach. First is the estimate of heterogeneity, which is an absolute assessment of heterogeneity. Second is the I^2 statistic, which is a relative assessment of heterogeneity, specifically the

Table 2
Unconscious Thought Theory Results

	Unconscious versus immediate	Unconscious versus distraction	Distraction versus immediate
(a) SCSM estimates			
Effect			
Estimate	0.5662	0.3928	0.2462
Standard Error	0.1010	0.2813	0.0943
z	5.6086	1.3964	2.6113
p	< 0.0001	0.1626	0.0090
Heterogeneity			
Estimate [95% CI]	0.1232 [0.0000, 0.4467]	0.3231 [0.1287, 2.2135]	0.0000 [0.0000, 0.5180]
I^2 [95% CI]	50.4631 [0.0000, 80.3013]	86.2326 [60.1673, 95.2416]	0.0000 [0.0000, 75.8445]
$Q(df;\, p)$	10.0935 (5; 0.0726)	14.5271 (2; 0.0007)	0.8613 (2; 0.6501)
(b) MCSM contrast and heterogeneity estimates			
Effect			
Estimate	0.5688	0.3482	0.2206
Standard Error	0.1132	0.1470	0.1470
z	5.0255	2.3681	1.5001
p	< 0.0001	0.0179	0.1336
Heterogeneity			
Estimate [95% CI]		0.1511 [0.0365, 0.4032]	
I^2 [95% CI]		59.4263 [11.5476, 81.3886]	
$Q(df;\, p)$		17.2526 (7; 0.0158)	
(c) MCSM variance-covariance matrix estimate			
		$\begin{pmatrix} 0.0128 & 0.0064 & 0.0064 \\ 0.0064 & 0.0216 & -0.0152 \\ 0.0064 & -0.0152 & 0.0216 \end{pmatrix}$	

proportion of the variation in the single-study estimates of an effect that is due to heterogeneity rather than sampling error; to place I^2 in context, Pigott (2012) defines low, medium, and high heterogeneity in psychological research as I^2 of 25%, 50%, and 75%, respectively. Third is the Q statistic, which is an assessment of the deviation from homogeneity, specifically a measure of the distance between the single-study estimates of an effect and the estimate from the basic fixed effects meta-analytic model which assumes homogeneity; to place Q in context, larger values indicate greater deviation from homogeneity as calibrated by the associated p-value, which employs a χ^2-distribution as the reference distribution under the null hypothesis of homogeneity. For more details, see, for example, Borenstein, Hedges, Higgins, and Rothstein (2009).

We begin by discussing the estimates of heterogeneity. We note that here and hereafter, all estimates of heterogeneity are presented on the same scale as the estimate of the effect (i.e., as a standard deviation rather than as variance) and normalized by the size of the associated contrast vector (i.e., divided by $\sqrt{\sum_i c_i^2}$ where the c_i denote the elements of the contrast vector) for comparability across effects. In this case study, the normalization factor is the same (and equal to $\sqrt{2}$) for all three effects because they are all simple effects; consequently, it is not necessary for comparability here. However, it will be important for comparability in the subsequent case studies where, for example, interaction effects are considered alongside simple effects. We note the scale factor of $\sqrt{2}$ means that the estimate of the heterogeneity of each effect is $\sqrt{2}$ times the estimate presented in the table.

As can be seen, the estimates of heterogeneity are highly inconsistent across the three effects: the point estimate based on the simple effect of unconscious thought versus immediate decision is 0.1232, with a 95% confidence interval estimate of [0.0000, 0.4467]; the point estimate based on the simple effect of unconscious thought versus mere distraction is 0.3231, with a 95% confidence interval estimate of [0.1287, 2.2135]; and the point estimate based on the simple effect of mere distraction versus immediate decision is 0.0000, with a 95% confidence interval estimate of [0.0000, 0.5180]. While it is possible that this inconsistency reflects differences in heterogeneity across the three effects, the inconsistency is not particularly diagnostic because estimates of heterogeneity from the basic random effects meta-analytic model have poor statistical properties when the number of observations (i.e., studies) is not large; for example, such estimates are biased, highly variable, highly inaccurate, and equal to zero implausibly often (Chung, Rabe-Hesketh, & Choi, 2013; Chung, Rabe-Hesketh, Dorie, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). Because there are only six observations of the simple effect of unconscious thought versus immediate decision and only three observations of both the simple effect of unconscious thought versus mere distraction and the simple effect of mere distraction versus immediate decision, such inconsistency would not be unexpected even were there no differences in heterogeneity across the three effects.

These facts regarding the statistical properties of estimates of heterogeneity have important implications for statistical inference. Specifically, the estimate of the standard error of an effect is a direct function of the estimate of heterogeneity: as the latter increases so too does the former. Thus, obtaining an accurate estimate of heterogeneity is important not only in and of itself but also for obtaining an accurate estimate of the standard error of an effect and thus also confidence interval estimates of and p-values for the effect (i.e., because these depend on the estimate of the standard error of the effect).

We now discuss the assessments of heterogeneity based on the I^2 statistic and the Q statistic. As can be seen, the assessments of heterogeneity are also highly inconsistent across the three effects. Specifically, the point estimate of I^2 based on the simple effect of unconscious thought versus immediate decision is 50.4631%, with a 95% confidence interval estimate of [0.0000%, 80.3013%]; the point estimate of I^2 based on the simple effect of unconscious thought versus mere distraction is 86.2326%, with a 95% confidence interval estimate of [60.1673%, 95.2416%]; and the point estimate of I^2 based on the simple effect of mere distraction versus immediate decision is 0.0000%, with a 95% confidence interval estimate of [0.0000%, 75.8445%]. Additionally, the p-value of the Q statistic for the simple effect of unconscious thought versus immediate decision is 0.0726; the p-value of the Q statistic for the simple effect of unconscious thought versus mere distraction is 0.0007; and the p-value of the Q statistic for the simple effect of mere distraction versus immediate decision is 0.6501. Again, this inconsistency is not particularly diagnostic because like estimates of heterogeneity from the basic random effects meta-analytic model, the assessments of heterogeneity based on the I^2 statistic and the Q statistic have poor statistical properties when the number of observations is not large; for example,

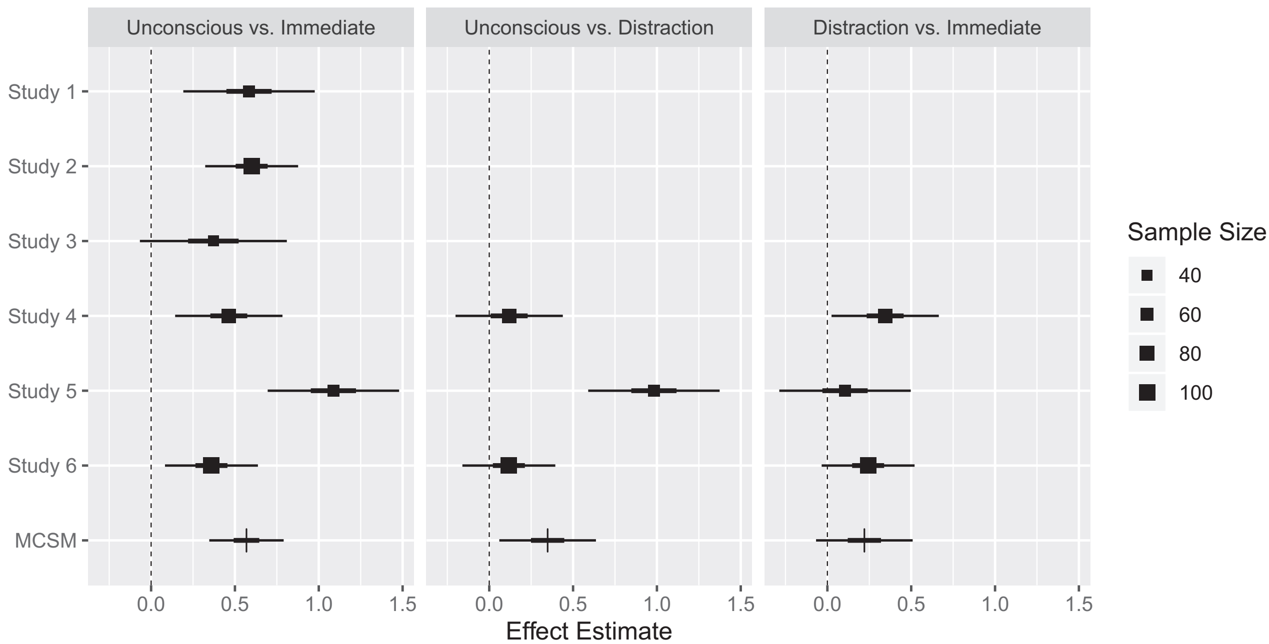

Figure 1. Unconscious Thought Theory Results. Point estimates are given by the squares for single-study estimates and the vertical bars for MCSM estimates; 50% and 95% confidence interval estimates are given by the thick and thin lines, respectively. The average sample size per condition in each study is given by the size of the squares.

the I^2 statistic is biased and both statistics suggest zero heterogeneity implausibly often (Higgins & Thompson, 2002; von Hippel, 2015; Huedo-Medina, Sánchez-Meca, Marın-Martınez, & Botella, 2006; Ioannidis, Patsopoulos, & Evangelou, 2007). Thus again, such inconsistency would not be unexpected even were there no differences in heterogeneity across the three effects.

We now proceed to discuss results from the MCSM approach, which can be found in the table and, alongside single-study results, in Figure 1. As can be seen, the estimates of the three effects are fully consistent with the predictions of unconscious thought theory. However, they are not particularly consistent with those from the SCSM approach. Specifically, while the point estimates of the three effects are unsurprisingly rather similar to those from the SCSM approach, the estimates of the standard errors differ considerably and, consequently, so too do, for example, the confidence interval estimates.

These differences in the estimates of the standard errors and thus the confidence interval estimates are in large part due to differences in the estimates of heterogeneity. Specifically, in contrast to the SCSM approach, the MCSM approach yields a single assessment of heterogeneity for all effects. The point estimate of heterogeneity is 0.1511, with a 95% confidence interval estimate of [0.0365, 0.4032]. Further, the point estimate of I^2 is 59.4263%, with a 95% confidence interval estimate of [11.5476%, 81.3886%] and the p-value of the Q statistic is 0.0158.

Finally, the MCSM approach yields an estimate of the covariance matrix of the estimates of the effects. Among other things, this allows for an assessment of the degree of dependence among them. For example, the estimate of the correlation between the estimates of the simple effect of unconscious thought versus mere distraction and the simple effect of mere distraction versus immediate decision is $-0.0152/\sqrt{0.0216 \cdot 0.0216} = -0.7038$, thereby suggesting the two estimates are highly dependent. The dependence is unsurprising, being driven in large part by the fact that the two effects both have the mere distraction condition in common. However, it is exacerbated by the mix of two-condition and three-condition study designs. For example, the simple effect of unconscious thought versus immediate decision and the simple effect of unconscious thought versus mere distraction both have the unconscious thought condition in common; similarly, the simple effect of unconscious thought versus immediate decision and the simple effect of mere distraction versus immediate decision both have the immediate decision condition in common. However, for these latter two pairs of effects, the correlation is substantially smaller in magnitude (i.e., 0.3849 for both pairs) because in both cases the common condition (i.e., unconscious thought and immediate decision, respectively) is included in all six studies whereas for the pair discussed in the beginning of this paragraph, the common condition (i.e., mere distraction) is included in only three studies.

In closing, we review how this case study illustrates several advantages that the MCSM approach has relative to the SCSM approach. As shown, it is user-friendly because it requires only basic summary information from each study condition like that found in Table 3. In contrast, the SCSM approach requires preprocessing of the data to convert effects to a standardized scale. Nonetheless,

Table 3
Choice Overload Hypothesis Study Summary Information

(a) Principal information

Study	Time pressure	Assortment size	Mean	Standard deviation	Sample size
Study 1	Absent	Small	3.8933	0.5829	75
	Absent	Large	3.7333	0.5022	75
	Present	Small			
	Present	Large			
Study 2	Absent	Small	6.8400	1.0514	100
	Absent	Large	6.5200	0.9154	100
	Present	Small			
	Present	Large			
Study 3	Absent	Small	65.9600	10.2598	50
	Absent	Large	69.3600	9.4820	50
	Present	Small			
	Present	Large			
Study 4	Absent	Small	70.0111	11.1239	90
	Absent	Large	73.8889	13.1955	90
	Present	Small	73.9000	11.2470	90
	Present	Large	65.9444	12.0694	90
Study 5	Absent	Small	6.7920	1.0341	125
	Absent	Large	6.7040	1.0162	125
	Present	Small	6.6800	1.0746	125
	Present	Large	6.3760	0.9727	125
Study 6	Absent	Small	6.6100	0.9200	100
	Absent	Large	6.7000	1.1849	100
	Present	Small	6.8100	0.9608	100
	Present	Large	6.4000	1.0731	100
Study 7	Absent	Small	3.8267	0.5526	150
	Absent	Large	3.8400	0.5319	150
	Present	Small	3.8400	0.5685	150
	Present	Large	3.6133	0.5768	150

(b) Study design and correlation information

Study	Time pressure	Assortment size	Time pressure	Assortment size	Correlation
Study 3	Absent	Small	Absent	Large	0.2206
Study 4	Absent	Small	Absent	Large	0.3204
	Present	Small	Present	Large	0.2219
Study 6	Absent	Small	Absent	Large	0.2159
	Absent	Small	Present	Small	0.0182
	Absent	Small	Present	Large	0.2108
	Absent	Large	Present	Small	0.4019
	Absent	Large	Present	Large	0.3495
	Present	Small	Present	Large	0.2018

both approaches yield results in standard deviation units in order to adjust for differences in the measurement scales used for the dependent measure across studies and so the results of the two approaches are directly comparable.

Additional advantages follow directly from the fact that the MCSM approach models the data from all conditions of all studies jointly in a single analysis whereas the SCSM approach models various subsets of the data in turn across multiple analyses. For example, in this case study, the MCSM approach models all fifteen observations (i.e., study conditions) jointly in a single analysis whereas the SCSM approach models six observations (i.e., one converted effect from each study) for the simple effect of unconscious thought versus immediate decision; separately models three observations (i.e., one converted effect each from Study 4, Study 5, and Study 6) for the simple effect of unconscious thought versus mere distraction; and separately models three observations (i.e., one converted effect each from Study 4, Study 5, and Study 6) for the simple effect of mere distraction versus immediate decision. This results in more accurate estimates of heterogeneity which in turn result in more accurate estimates of the standard errors of the estimates of the effects and therefore improved statistical inference. It also allows for the estimation of the covariance matrix of the estimates of the effects which is not possible with separate analyses.

Additionally, the MCSM approach yields coherent estimates of the effects while the SCSM approach does not. For example, in this case study, the simple effect of mere distraction versus immediate decision is by definition equal to the simple effect of unconscious thought versus immediate decision minus the simple effect of unconscious thought versus mere distraction (i.e., (Distraction − Immediate) = (Unconscious − Immediate) − (Unconscious − Distraction)). This is reflected in, for example, the point estimates obtained via the MCSM approach (i.e., $0.5688 - 0.3482 = 0.2206$) but it is not reflected in the point estimates obtained via the SCSM approach (i.e., $0.5662 - 0.3928 = 0.1734 \neq 0.2462$).

A final advantage is illustrated in part by the fact that the MCSM approach accommodates a mix of study designs, here both two-condition and three-condition study designs. In contrast, the SCSM approach here in effect treats each simple contrast as if it came from a two-condition study featuring the relevant two conditions and ignoring any others.

Given these advantages of the MCSM approach relative to the SCSM approach, we focus only on the MCSM approach in the next two case studies.

Case Study II: Choice Overload Hypothesis

A researcher was interested in conducting a meta-analysis of studies of the choice overload hypothesis, the conjecture that an increase in the number of options from which to choose can result in adverse consequences such as a decrease in the likelihood of making a choice or a decrease in the satisfaction with a choice. The researcher was particularly interested in the moderating role of time pressure. Toward this end, the researcher surveyed the literature and identified as relevant seven hypothetical studies.

Three of the studies followed a two-condition study design (small or large assortment size with time pressure absent) while the other four followed a two-by-two study design (small or large assortment size; time pressure absent or present). Further, four of the studies followed a between-subjects study design while the other three followed a within-subjects study design; specifically, Study 3 and Study 6 followed a fully within-subjects study design and Study 4 followed a partially within-subjects study design with assortment size as a within-subjects factor and time pressure as a between-subjects factor. Finally, the dependent measure in these studies was satisfaction which was measured on a five-point integer scale in Study 1 and Study 7; a nine-point integer scale in Study 2, Study 5, and Study 6; and a one hundred point integer scale in Study 3 and Study 4.

The effects of interest were (a) the simple effect of assortment size when time pressure was absent, (b) the simple effect of assortment size when time pressure was present, and (c) the interaction effect, which are given by the respective contrast vectors $(-1\ 1\ 0\ 0)$, $(0\ 0\ -1\ 1)$, and $(1\ -1\ -1\ 1)$. The first effect was predicted to be null while the other two effects were predicted to be negative.

Summary information for these studies can be found in Table 3. As illustrated in the table, when one or more studies follows a within-subjects study design as here, information on the study designs (specifically, the identity of the within-subjects study condition pairs) and the correlation of the individual-level observations between these pairs may optionally be provided in which case the MCSM approach will account for it.

Results from the MCSM approach can be found in Table 4 and, alongside single-study results, in Figure 2. As can be seen, the estimates of the three effects are fully consistent with the predictions of the choice overload hypothesis. Further, the point estimate of heterogeneity is 0.1029, with a 95%

Table 4
Choice Overload Hypothesis MCSM Results

	Simple Effect 1	Simple Effect 2	Interaction Effect
Effect			
Estimate	0.0109	−0.4363	−0.4472
Standard Error	0.0756	0.0958	0.1228
z	0.1441	−4.5527	−3.6418
p	0.8855	<0.0001	0.0003
Heterogeneity			
Estimate [95% CI]	0.1029 [0.0186, 0.2050]		
I^2 [95% CI]	55.0531 [16.1049, 75.9197]		
$Q(df;\ p)$	26.6982 (12; 0.0085)		

Note. Simple Effect 1 denotes the simple effect of assortment size when time pressure was absent; Simple Effect 2 denotes the simple effect of assortment size when time pressure was present; and Interaction Effect denotes the assortment size × time pressure interaction effect.

confidence interval estimate of [0.0186, 0.2050]. Finally, the point estimate of I^2 is 55.0531%, with a 95% confidence interval estimate of [16.1049%, 75.9197%] and the p-value of the Q statistic is 0.0085.

One potential objection to these results is that researchers may not have access to one or more of the correlations that can be found in Table 3 (i.e., because the correlations were not reported in a primary analysis and the researchers do not possess the individual-level observations from which they can be computed) and that are required for these results. While one could then treat all studies as following a between-subjects study design, we note that one can do better with the MCSM approach. Specifically, even when a researcher *does not* have access to one or more of these correlations, the researcher typically *does* have access to the study designs, specifically here that Study 3 and Study 6 followed a fully within-subjects study design and Study 4 followed a partially within-subjects study design with assortment size as a within-subjects factor and time pressure as a between-subjects factor. Consequently, the MCSM approach can make use of this study design information to at least in part account for the dependence among the relevant within-subjects study condition pairs (in particular, via the estimates of heterogeneity), even when assuming the unknown correlations are equal to, for example, zero.

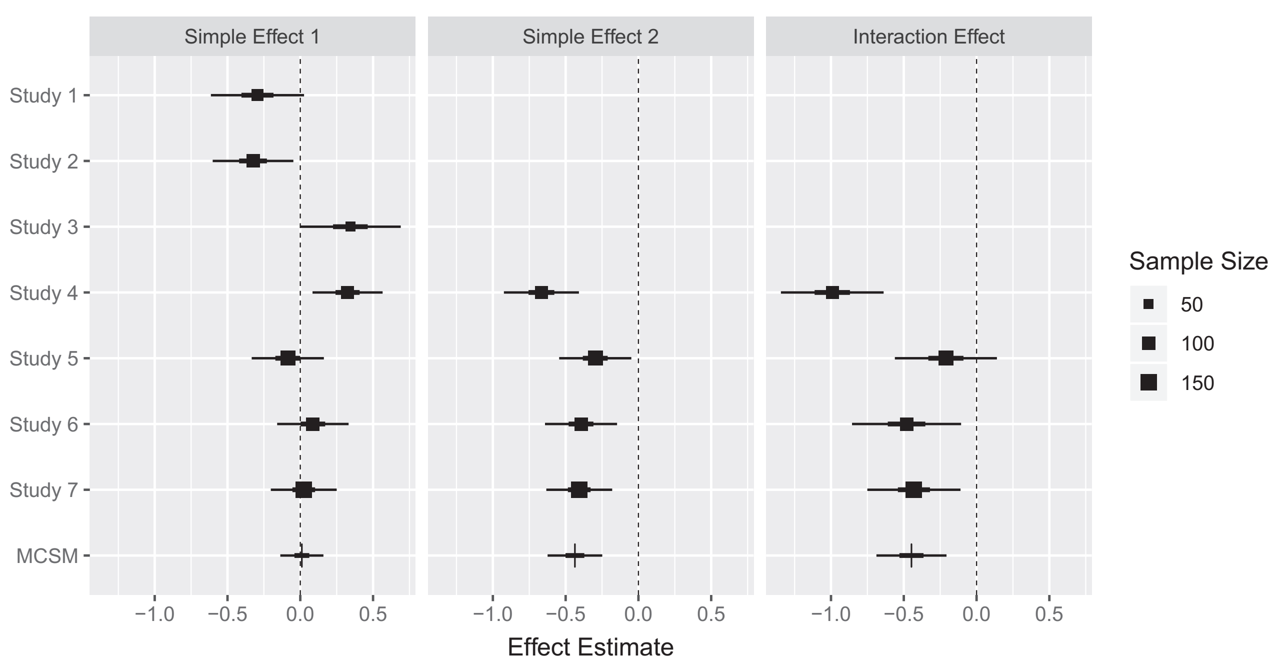

Figure 2. Choice Overload Hypothesis Results. Point estimates are given by the squares for single-study estimates and the vertical bars for MCSM estimates; 50% and 95% confidence interval estimates are given by the thick and thin lines, respectively. The average sample size per condition in each study is given by the size of the squares. Simple Effect 1 denotes the simple effect of assortment size when time pressure was absent; Simple Effect 2 denotes the simple effect of assortment size when time pressure was present; and Interaction Effect denotes the assortment size × time pressure interaction effect.

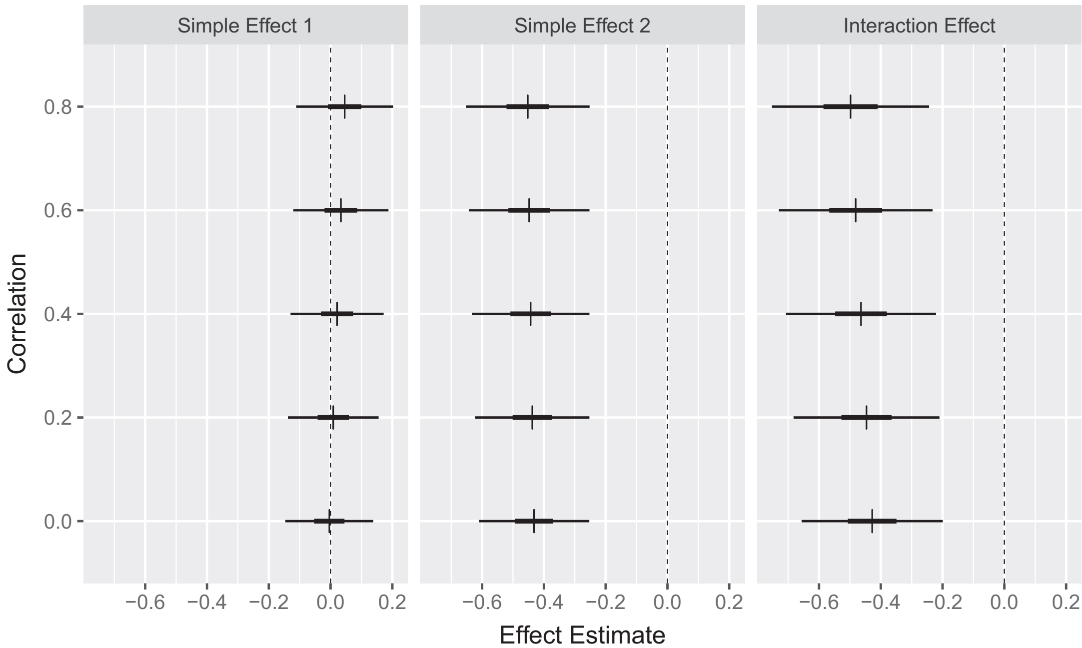

Figure 3. Choice Overload Hypothesis MCSM Sensitivity Analysis. Point estimates are given by the vertical bars; 50% and 95% confidence interval estimates are given by the thick and thin lines, respectively. Simple Effect 1 denotes the simple effect of assortment size when time pressure was absent; Simple Effect 2 denotes the simple effect of assortment size when time pressure was present; and Interaction Effect denotes the assortment size × time pressure interaction effect.

Further, the MCSM approach allows for a sensitivity analysis for the unknown correlations, in particular by setting them all equal to various successive values (i.e., 0.0, 0.2, 0.4, 0.6, and 0.8) in turn. Such a sensitivity analysis can be found in Figure 3. As can be seen, the results are not sensitive to the value of the correlation in this case study.

Case Study III: Flexible Correction Model

A researcher was interested in conducting a meta-analysis of studies of the flexible correction model, the conjecture according to which individuals will correct for a bias in evaluating a source of information when they are motivated and able to search for potential sources of bias; find one or more such sources; have or are able to generate a theory regarding the direction and magnitude of the bias resulting from the source(s); and are motivated and able to correct for the bias based on this theory. The researcher was particularly interested in the moderating role of the likability of the source of information. Toward this end, the researcher surveyed the literature and identified as relevant ten hypothetical studies.

All of the studies followed a two-by-two study design (likable or unlikable source; correction instructions absent or present). Further, all of the studies followed a between-subjects study design. Finally, the dependent measure in these studies was the persuasiveness of the information which was measured on a five-point integer scale in Study 1 and Study 6; a nine-point integer scale in Study 2, Study 4, Study 5, and Study 10; an eleven-point integer scale in Study 8 and Study 9; and a twenty-point integer scale in Study 3 and Study 7.

The effects of interest were (a) the simple effect of the correction instruction when the source was likable, (b) the simple effect of the correction instruction when the source was unlikable, and (c) the interaction effect, which are given by the respective contrast vectors $(-1\ 1\ 0\ 0)$, $(0\ 0\ -1\ 1)$, and $(1\ -1\ -1\ 1)$. The first effect was predicted to be negative while the other two effects were predicted to be positive.

The researcher was also interested in two study-level moderators. First, when feelings toward a source are very strong and salient (e.g., when a source is extremely likable or unlikable), subjects sometimes report ignoring the correction instructions (perhaps because they are unable to correct for bias even when instructed to do so; see footnotes 4–6 of Petty et al., (1998)). Consequently, the researcher coded whether or not subjects reported ignoring the correction instructions. Second, some hold that scientific claims receive decreasing support over time, the so-called "decline effect." Consequently, the researcher coded the year the study was conducted.

Summary information for these studies can be found in Table 5. As illustrated in the table, when one or more study-level moderators are of interest as here, information on the moderators (specifically, their value for each study) may be provided in which case the MCSM approach will account for it.

We begin by discussing results ignoring study-level moderators. Results from the MCSM approach ignoring study-level moderators can be found in Table 6 and, alongside single-study results, in Figure 4. As can be seen, the estimates of the three effects are not fully consistent with the predictions of the flexible correction model. Further, the point estimate of heterogeneity is 0.2794, with a 95% confidence interval estimate of [0.2073, 0.3975]. Finally, the point estimate of I^2 is 84.4518%, with a 95%

Table 5
Flexible Correction Model Study Summary Information

Study	Likability	Correction instructions	Mean	Standard deviation	Sample size	Instructions ignored	Year
Study 1	Likable	Absent	3.1200	0.5584	50	0	2000
	Likable	Present	3.0200	0.6224	50		
	Unlikable	Absent	2.8600	0.4953	50		
	Unlikable	Present	2.9000	0.4629	50		
Study 2	Likable	Absent	5.4000	1.0000	75	0	2001
	Likable	Present	5.0667	1.0696	75		
	Unlikable	Absent	4.5467	1.0941	75		
	Unlikable	Present	4.7200	1.2031	75		
Study 3	Likable	Absent	11.9200	2.3254	100	1	2002
	Likable	Present	12.0500	2.3198	100		
	Unlikable	Absent	9.0100	2.2360	100		
	Unlikable	Present	8.6600	2.2029	100		
Study 4	Likable	Absent	5.2500	1.0316	40	0	2003
	Likable	Present	4.9750	1.0250	40		
	Unlikable	Absent	4.4250	1.0099	40		
	Unlikable	Present	4.9500	0.7828	40		
Study 5	Likable	Absent	5.3833	1.0750	60	0	2006
	Likable	Present	4.8333	0.8862	60		
	Unlikable	Absent	4.7333	1.0229	60		
	Unlikable	Present	5.3000	0.8497	60		
Study 6	Likable	Absent	3.2300	0.5096	100	0	2010
	Likable	Present	2.9900	0.5773	100		
	Unlikable	Absent	2.8600	0.6034	100		
	Unlikable	Present	2.9500	0.5752	100		
Study 7	Likable	Absent	12.1200	2.6699	50	1	2012
	Likable	Present	12.1000	2.3669	50		
	Unlikable	Absent	9.2200	2.2883	50		
	Unlikable	Present	8.3600	2.1831	50		
Study 8	Likable	Absent	6.5400	1.4875	50	1	2013
	Likable	Present	6.8600	1.4848	50		
	Unlikable	Absent	5.5200	1.2329	50		
	Unlikable	Present	5.4000	1.3702	50		
Study 9	Likable	Absent	6.6000	1.3457	75	0	2016
	Likable	Present	5.8667	1.2339	75		
	Unlikable	Absent	5.2533	1.3058	75		
	Unlikable	Present	5.5867	1.1751	75		
Study 10	Likable	Absent	5.5200	1.0297	100	1	2018
	Likable	Present	5.6300	1.0016	100		
	Unlikable	Absent	4.5600	1.0854	100		
	Unlikable	Present	4.5200	1.0198	100		

Table 6
Flexible Correction Model MCSM Results Ignoring Study-level Moderators

	Simple Effect 1	Simple Effect 2	Interaction Effect
Effect			
Estimate	−0.1970	0.1115	0.3085
Standard Error	0.1368	0.1368	0.1935
z	−1.4400	0.8145	1.5942
p	0.1499	0.4153	0.1109
Heterogeneity			
Estimate [95% CI]	0.2794 [0.2073, 0.3975]		
I^2 [95% CI]	84.4518 [78.5618, 88.7236]		
$Q(df;\ p)$	173.6540 (27; $<$0.0001)		

Note. Simple Effect 1 denotes the simple effect of the correction instruction when the source was likable; Simple Effect 2 denotes the simple effect of the correction instruction when the source was unlikable; and Interaction Effect denotes the source × correction instructions interaction effect.

confidence interval estimate of [78.5618%, 88.7236%] and the p-value of the Q statistic is less than 0.0001. This degree of heterogeneity is consistent with the fact that the single-study results (which can be found in the figure) vary considerably from study to study and suggests unaccounted for study-level moderators.

Given this, we proceed to discuss results including study-level moderators. Results from the MCSM approach accounting for the study-level moderators can be found in Table 7. The estimates of the intercepts give, as always, estimates of the effects when all study-level moderators are set to zero. Because this would indicate the estimates when the correction instructions are heeded and the year is zero and because this is not particularly meaningful given that all studies were conducted between 2000 and 2018, we subtracted 2000 from the year in the analysis. Consequently, the estimates of the intercepts give the estimates of the effects when correction instructions are heeded and the year is 2000 (i.e., the year the first study was conducted). As can be seen, these estimates are fully consistent with the predictions of the flexible correction model. On the other hand, the estimates of the three coefficients for the study-level moderator indicating that subjects ignored the correction instructions are all opposite in sign to and larger in magnitude than those of the intercepts; this suggests that the effects will be attenuated and potentially even reversed when subjects ignored the correction instructions. In addition, the estimates of the three coefficients for the study-level moderator indicating the year the study was conducted are all negligible in

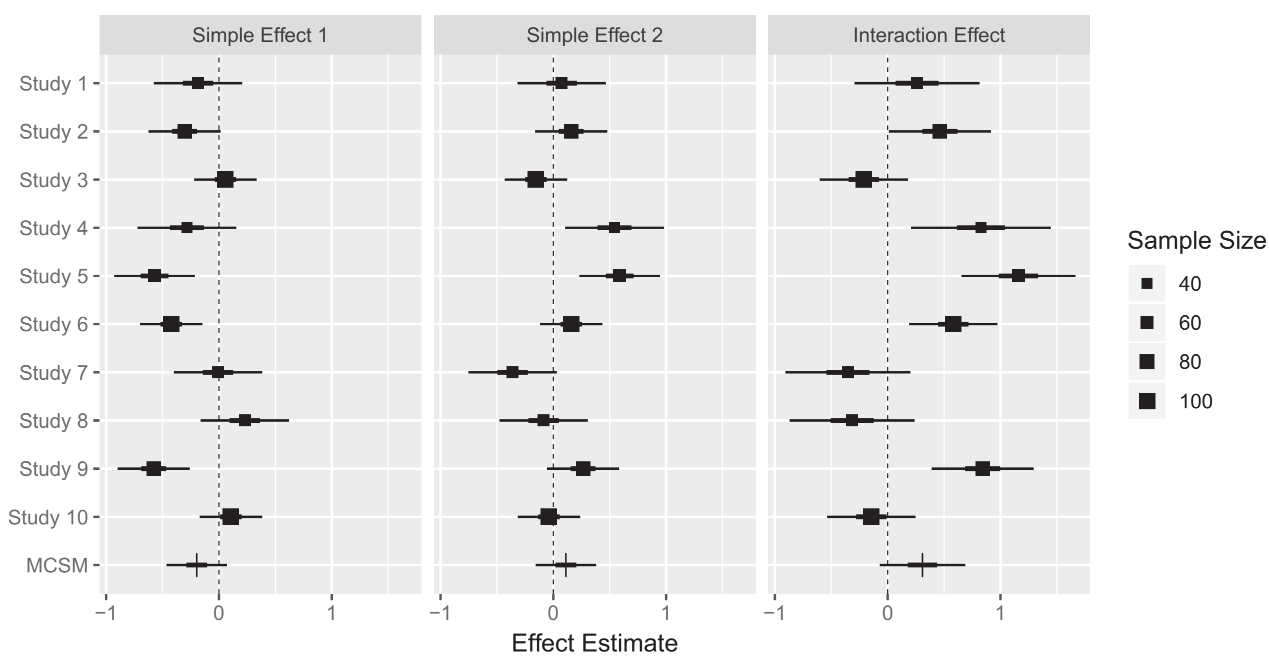

Figure 4. Flexible Correction Model Results Ignoring Study-level Moderators. Point estimates are given by the squares for single-study estimates and the vertical bars for MCSM estimates; 50% and 95% confidence interval estimates are given by the thick and thin lines, respectively. The average sample size per condition in each study is given by the size of the squares. Simple Effect 1 denotes the simple effect of the correction instruction when the source was likable; Simple Effect 2 denotes the simple effect of the correction instruction when the source was unlikable; and Interaction Effect denotes the source × correction instructions interaction effect.

Table 7
Flexible Correction Model MCSM Results Including Study-level Moderators

	Simple Effect 1	Simple Effect 2	Interaction Effect
Intercept			
Estimate	−0.3435	0.2580	0.6015
Standard Error	0.1229	0.1229	0.1738
z	−2.7950	2.0990	3.4605
p	0.0052	0.0358	0.0005
Instructions ignored			
Estimate	0.5364	−0.4452	−0.9816
Standard Error	0.1559	0.1559	0.2205
z	3.4402	−2.8554	−4.4517
p	0.0006	0.0043	<0.0001
Year			
Estimate	−0.0090	0.0036	0.0126
Standard Error	0.0121	0.0121	0.0171
z	−0.7460	0.2967	0.7373
p	0.4557	0.7667	0.4609
Heterogeneity			
Estimate [95% CI]	0.1048 [0.0172, 0.1997]		
I^2 [95% CI]	41.8822 [3.5048, 64.9964]		
$Q(df;\, p)$	36.1335 (21; 0.0211)		

Note. Simple Effect 1 denotes the simple effect of the correction instruction when the source was likable; Simple Effect 2 denotes the simple effect of the correction instruction when the source was unlikable; and Interaction Effect denotes the source × correction instructions interaction effect.

magnitude, suggesting that the year the study was conducted is not substantially associated with the results. Finally, including study-level moderators substantially reduces heterogeneity.

It can often be more meaningful and interpretable to consider estimates of the effects at specific study-level moderator levels rather than estimates of intercepts and coefficients as above. Results from the MCSM approach when the correction instructions are heeded versus ignored and the year is set, respectively, to 2000, 2006, 2012, and 2018 can be found in Figure 5. As can be seen, these estimates are fully consistent with the predictions of the flexible correction model when the correction instructions manipulation is successful but they are not when it is not. Further, the year the study was conducted is not substantially associated with the estimates.

Discussion

In this section, we discuss some cautions relevant when applying the MCSM approach. We note that these cautions apply not only to the MCSM approach but also to the SCSM approach which the MCSM approach generalizes as well as to many other meta-analytic approaches. We also note that these cautions are not criticisms. We conclude with a brief summation.

Adjustment for Differences in Measurement Scales

The MCSM approach, the SCSM approach, and many other meta-analytic approaches yield results on a common standardized scale in order to adjust for differences in the measurement scales used for the dependent measure across studies. Caution is due when applying these approaches. Specifically, one should be careful that the same or a similar construct is assessed across studies and that the differences are solely in the measurement scales. Moreover, one should be careful that conversion to a common standardized scale is reasonable, in particular, that any and all differences in how individuals might respond to differences in the measurement scales are accounted for by the standardization.

In addition, one should be aware that while these approaches are advantageous in that they can accommodate a great variety of studies (i.e., because the measurement scales used for the dependent measure across studies need not be the same), they come with a major limitation when the measurement scale used for the dependent measure across studies is in fact the same and thus such adjustment is unnecessary. Specifically, such approaches can account only for heterogeneity that is idiosyncratic to the conditions within a given study and cannot account for any heterogeneity that is common across them; in other words, heterogeneity involving differences in levels across studies is not identified and only heterogeneity involving differences in contrasts across studies is identified. In contrast, approaches that do not adjust for differences in the measurement scales used for the dependent measure across studies (e.g., McShane & Böckenholt, 2017) can account for both differences in levels and differences in contrasts; consequently, these approaches are preferable when applicable (i.e., when the measurement scale used for the dependent measure across studies is the same).

To illustrate the distinction between heterogeneity involving differences in levels and heterogeneity involving differences in contrasts, consider the three study scenarios depicted in Figure 6 in which the measurement scales used for the dependent

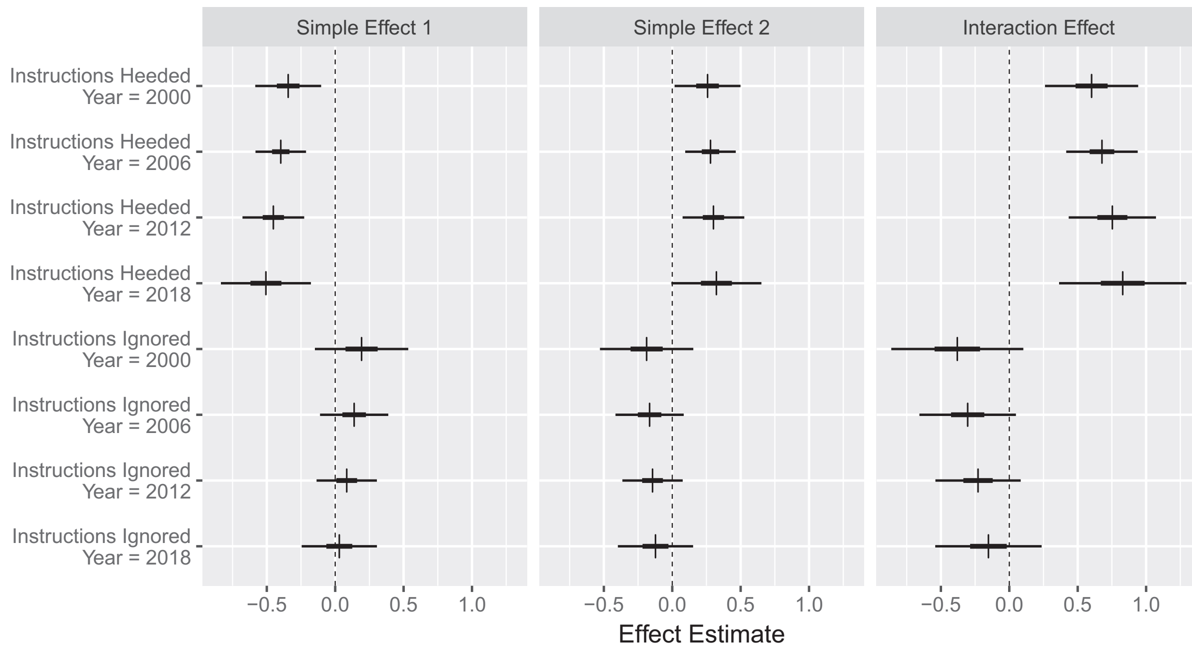

Figure 5. Flexible Correction Model MCSM Results at Eight Study-level Moderator Levels. Point estimates are given by the vertical bars; 50% and 95% confidence interval estimates are given by the thick and thin lines, respectively. Simple Effect 1 denotes the simple effect of the correction instruction when the source was likable; Simple Effect 2 denotes the simple effect of the correction instruction when the source was unlikable; and Interaction Effect denotes the source × correction instructions interaction effect.

measure across the studies is the same five-point integer scale and in which, for simplicity, all studies follow the same two-condition study design. We note the hypothetical data used in these three study scenarios is not meant to be realistic but rather is meant to clearly demonstrate this distinction.

In the first scenario, the studies differ in terms of levels but not in terms of contrasts (i.e., here the simple contrast between the two conditions); put differently, there is heterogeneity that is common across the conditions within a given study but there is no heterogeneity that is idiosyncratic to each condition. In the second scenario, the studies differ in terms of contrasts but not in terms of levels; put differently, there is no heterogeneity that is common across the conditions within a given study but there is heterogeneity that is idiosyncratic to each condition. In the third scenario, the studies differ both in terms of levels and in terms contrasts; put differently, there is both heterogeneity that is common across the conditions within a given study and heterogeneity that is idiosyncratic to each condition.

Consequently, approaches that adjust for differences in the measurement scales used for the dependent measure across studies cannot account for and therefore detect no heterogeneity in the first scenario; fully account for heterogeneity in the second scenario; and only partially account for heterogeneity in the third scenario (i.e., accounting only for the heterogeneity that is idiosyncratic to each condition or that involving differences in contrasts). In contrast, approaches that do not adjust for differences in the measurement scales used for the dependent measure across studies fully account for heterogeneity in all three scenarios.

For these reasons, it is preferable in meta-analysis, as in statistical analysis more generally, to use approaches which yield results on the original measurement scale when possible (e.g., when the measurement scale used for the dependent measure across studies is the same; Baguley (2009), Bond, Wiitala, and Richard (2003), Greenland, Schlesselman, and Criqui (1986), Tukey (1969), and Wilkinson (1999)).

Relationship Among Studies

The MCSM approach, the SCSM approach, and many other meta-analytic approaches assume (conditional) independence of the studies included in the meta-analysis. This amounts to assuming each study has the same relationship to each other study. This assumption may be tenable for meta-analyses of the small number of studies from a single paper or

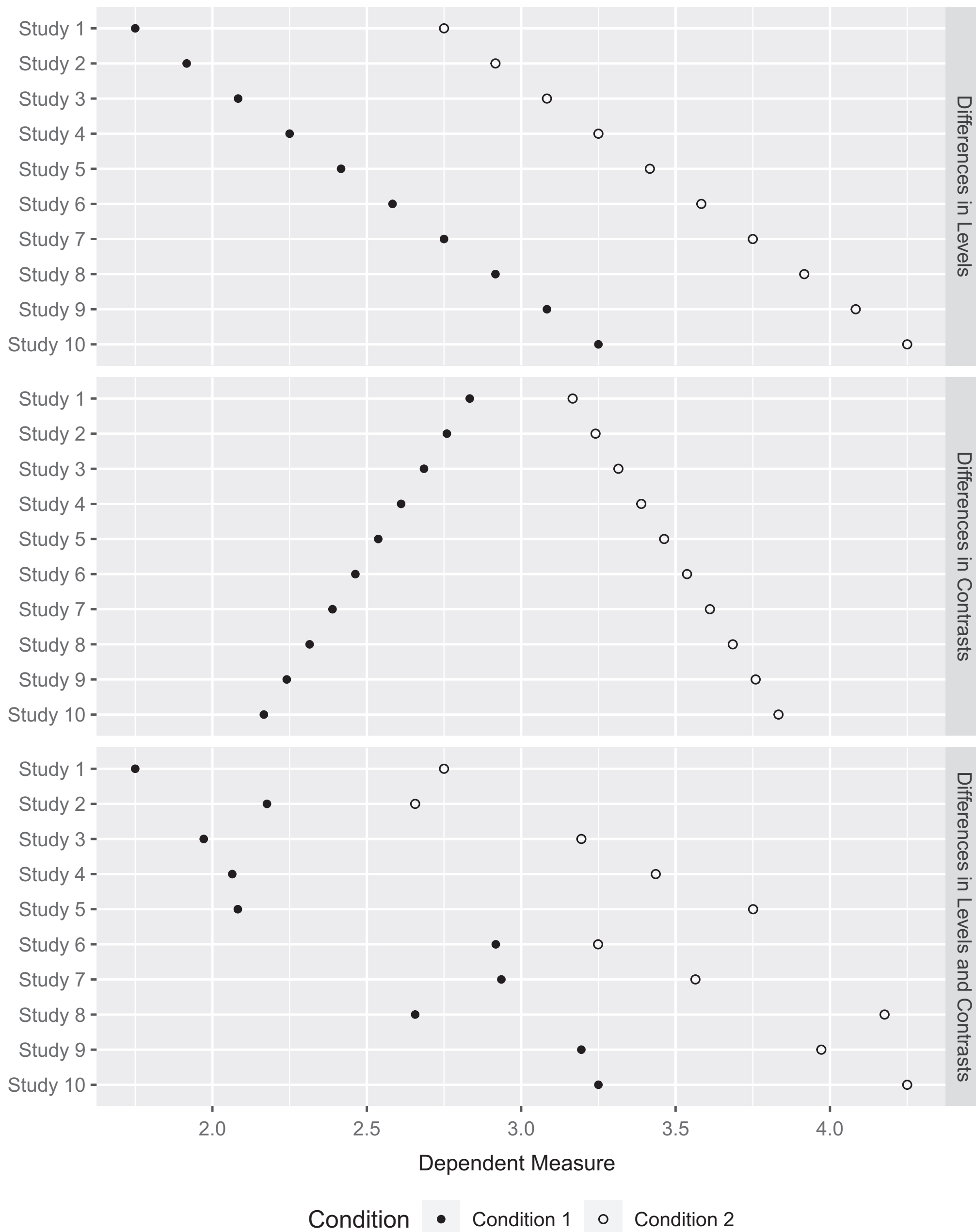

Figure 6. Three Study Scenarios. The means of the dependent measure in each condition of each study are given by the points.

meta-analyses of a large number of studies from multiple papers in which each paper contributes a single study. However, it is less tenable for, for example, meta-analyses of a large number of studies from multiple papers in which each paper contributes multiple studies because studies from the same paper will typically have a closer relationship to one another than studies from different papers. In practice, multilevel multivariate meta-analytic approaches are typically required for such meta-analyses (Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, & Colditz, 1998; Cheung, 2015; Kalaian & Raudenbush, 1996; McShane & Böckenholt, 2018).

Study-level Moderators

The MCSM approach, the SCSM approach, and many other meta-analytic approaches accommodate study-level moderators. Caution is due when interpreting estimates of the effects of study-level moderators. Specifically, one cannot in general interpret these estimates as causal because such moderators are typically observed and not manipulated. Therefore, one can in general interpret these estimates as only associational. Nonetheless, these estimates present an opportunity for validation in future research.

Summation

In this paper, we have introduced novel meta-analytic methodology that generalizes the SCSM approach to meta-analysis to accommodate an arbitrary number of contrasts. This MCSM approach has several advantages relative to the SCSM approach: It is user-friendly, yields more accurate estimates of heterogeneity which in turn results in more accurate estimates of the standard errors of the estimates of the effects and therefore improved statistical inference, yields estimates of the covariance matrix of the estimates of the effects, yields coherent estimates of the effects, and accommodates a mix of study designs and study-level moderators.

Although we have focused on contrasts of the means of the multiple study conditions that arise from the variation of one or more experimental factors, the MCSM approach is fully general in that it can accommodate any type of contrasts of means. For example, in observational research with no study conditions, contrasts of the means of the multiple subgroups of subjects (e.g., demographic groups) that arise from the variation of one or more discrete individual-level covariates are often of interest. In such cases, the MCSM approach can accommodate this by allowing those subgroups of subjects to play the role of the study conditions.

In closing, meta-analysis is a powerful tool for the estimation of effects, heterogeneity, and the association of study-level moderators. We are optimistic that the MCSM approach will prove useful for the meta-analysis of studies with multiple contrasts and differences in the measurement scales used for the dependent measure across studies.

References

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, *17*(22), 2537–2550.

Bond, C. F. Jr, Wiitala, W. L., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods*, *8*(4), 406.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Bos, M. W., Dijksterhuis, A., & van Baaren, R. B. (2011). The benefits of "sleeping on things": Unconscious thought leads to automatic weighting. *Journal of Consumer Psychology*, *21*, 4–8.

Cheung, M.-W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. Chichester, UK: Wiley.

Chung, Y., Rabe-Hesketh, S., & Choi, I.-H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine*, *32*(23), 4071–4089.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A non-degenerate estimator for hierarchical variance parameters via penalized likelihood estimation. *Psychometrika*, *78*(4), 685–709.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, *14* (2), 165–176.

Greenland, S., Schlesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology*, *123*(2), 203–208.

Haidich, A. (2010). Meta-analysis in medical research. *Hippokratia*, *14*(1), 29–37.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72* (358), 320–338.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558.

Huedo-Medina, T. B., Sánchez-Meca, J., Marın-Martınez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or i2 index? *Psychological Methods*, *11*(2), 193.

Ioannidis, J. P., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *BMJ: British Medical Journal*, *335*(7626), 914.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, *1*(3), 227–235.

McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability. *Journal of Consumer Research*, *43*(6), 1048–1063.

McShane, B. B., & Böckenholt, U. (2018). Multilevel multivariate meta-analysis with application to choice overload. *Psychometrika*, *83*(1), 255–271.

McShane, B. B., & Böckenholt, U. (2020). Enriching meta-analytic models of summary data: A thought experiment and case study. *Advances in Methods and Practices in Psychological Science*, *3*(1), 81–93.

Petty, R. E., Wegener, D. T., & White, P. H. (1998). Flexible correction processes in social judgment: Implications for persuasion. *Social Cognition*, *16*(1), 93–113.
Pigott, T. (2012). *Advances in meta-analysis*. New York, NY: Springer.
Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, *6*(1), 15–32.
Simmonds, M. C., Higgins, J. P., Stewart, L. A., Tierney, J. F., Clarke, M. J., & Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. *Clinical Trials*, *2*(3), 209–217.
Stewart, L. A., & Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions*, *25*(1), 76–97.
Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*(2), 83–91.
von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1), 35.
Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604.

Appendix A: Model Specification and Estimation Procedure

In this appendix, we discuss the model specification and estimation procedure underlying the MCSM approach. We initially assume, for notational simplicity, that each study follows a between-subjects study design and study-level moderators are not of interest; we later relax these assumptions. We then demonstrate the equivalence of the MCSM approach and the SCSM approach when all studies follow the same two-condition study design and interest centers on the simple contrast between the two conditions as measured on the standardized mean difference scale. We conclude by discussing how we can further relax our assumptions.

We let y_i, σ_i, and n_i denote the respective mean, standard deviation, and sample size of the individual-level observations in some study condition i and let $s[i]$ and $c[i]$ denote the respective study and condition to which these observations belong. Our model specification for the y_i is given by

$$y_i = \sigma_{s[i]}\left(\alpha_{s[i]} + \beta_{c[i]} + \gamma_i + \varepsilon_i\right)$$

where σ_s denotes the pooled standard deviation of study s; the α_s are treated as fixed effects that model the location of each study; the β_c are treated as fixed effects that model each condition with β_1 set to zero for identifiability; the γ_i are treated as random effects for each study condition; and the ε_i are random errors for each study condition. We further assume that the γ_i are independent and identically normally distributed with mean zero and variance τ^2; the ε_i are independent and normally distributed with mean zero and variance $1/n_i$; and there is zero covariation among the γ_i and ε_j.

We now relax our simplifying assumptions that each study follows a between-subjects study design and study-level moderators are not of interest. When one or more studies follows a within-subjects study design, our model specification for the y_i is given by

$$y_i = \sigma_{s[i]}\left(\alpha_{s[i]} + \beta_{c[i]} + \gamma_i + \delta_{g[i]} + \varepsilon_i\right)$$

where $g[i]$ denotes the group of subjects featured in study condition i, the δ are treated as random effects for each group of subjects, and everything else remains as above. Letting $\boldsymbol{\varepsilon}$ denote the vector of the ε_i, we further assume that the δ_g are independent and identically normally distributed with mean zero and variance ω^2; $\boldsymbol{\varepsilon}$ is multivariate normally distributed with mean and variance as above and covariance of zero when study conditions i and j feature a distinct group of subjects (i.e., $g[i] \neq g[j]$) and $\rho_{i,j}/n_i$ when study conditions i and j feature the same group of subjects (i.e., $g[i] = g[j]$) where necessarily $n_i = n_j$ and where $\rho_{i,j}$ is the correlation between the individual-level observations in study conditions i and j; and there is zero covariation among the γ_i, δ_g, and ε_j.

When study-level moderators are of interest, our model specification remains as above but the β_c are replaced by β_i which are parameterized as

$$\beta_i = \beta_{0,c[i]} + \beta_{1,c[i]} x_{1,s[i]} + \ldots + \beta_{p,c[i]} x_{p,s[i]}$$

where $x_{j,s}$ is the value of moderator j in study s and the $\beta_{j,c}$ are treated as fixed effects with the $\beta_{j,1}$ set to zero for identifiability.

Estimation of our model is as follows. We first estimate the σ_s (and, when one or more studies follows a within-subjects study design, the $\rho_{i,j}$) using conventional formulae; as these play a role in our model specification akin to that played by the sampling variance in standard meta-analysis, we follow the typical meta-analytic practice of treating them as known. Next, we estimate τ (or, when one or more studies follows a within-subjects study design, τ and ω) using restricted (or residual or reduced) maximum likelihood conditional on the estimates of the σ_s (Harville, 1977). Finally, we estimate the α_s, the β_c (or, when study-level moderators are of interest, the $\beta_{j,c}$), and their variance-covariance matrix using the standard best linear unbiased prediction estimators conditional on the estimates of the σ_s and τ (or, when one or more studies follows a within-subjects study design, the σ_s, the $\rho_{i,j}$, τ, and ω; Robinson (1991)); estimates of contrasts follow directly from these.

We now demonstrate the equivalence of the MCSM approach and the SCSM approach when all studies follow the same two-condition between-subjects study design and interest centers on the simple contrast between the two conditions as measured on the standardized mean difference scale. Let i_1 and i_2 index the two study conditions of a given study so that $s[i_1] = s[i_2] = s$, $c[i_1] = 1$, and $c[i_2] = 2$. Take the difference between our model specification equations for the two study conditions and divide by σ_s to yield

$$\frac{y_{i_2} - y_{i_1}}{\sigma_s} = (\beta_2 - \beta_1) + \left(\gamma_{i_2} - \gamma_{i_1}\right) + (\varepsilon_{i_2} - \varepsilon_{i_1}).$$

The left hand side of this equation is the standardized mean difference for the study, and, by the above, $(\beta_2 - \beta_1) = \beta_2$ is treated as a fixed effect that models the difference between the two conditions; $\left(\gamma_{i_2} - \gamma_{i_1}\right)$ is treated as a random effect for the study and across studies these are independent and identically normally distributed with mean zero and variance $2\tau^2$; $(\varepsilon_{i_2} - \varepsilon_{i_1})$ is treated as a random error for the study and across studies these are independent and normally distributed with mean zero and variance $1/n_{i_1} + 1/n_{i_2}$; and there is zero covariation among the $\left(\gamma_{i_2} - \gamma_{i_1}\right)$ and $\left(\varepsilon_{j_2} - \varepsilon_{j_1}\right)$. This is the model specification of the SCSM approach. The same holds *mutatis mutandis* when all studies follow the same two-condition within-subjects study design and/or study-level moderators are of interest.

We note we can relax our assumptions on the γ_i. Specifically, letting $\boldsymbol{\gamma}_s$ denote the vector of the γ_i for each study, we can instead assume the $\boldsymbol{\gamma}_s$ are independent and identically multivariate normally distributed with mean zero and variance-covariance matrix $\mathbf{T}$ where τ_{11}—and thus the τ_{1j} and τ_{j1}—are set to zero for identifiability and where τ_{jk} denotes the jk^{th} element of $\mathbf{T}$. The same holds *mutatis mutandis* regarding our assumptions on the δ_g when one or more studies follows a within-subjects study design.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Methodological Details Appendix.