View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Communications of the Association for Information Systems

Volume 42 Article 8

2-2018

Ranking Factors by Importance in Factorial Survey Analysis

Robert F. Otondo
Mississippi State University, bob.otondo@msstate.edu

Robert E. Crossler
Washington State University

Merrill Warkentin
Mississippi State University

Follow this and additional works at: http://aisel.aisnet.org/cais

Recommended Citation
Otondo, Robert F.; Crossler, Robert E.; and Warkentin, Merrill (2018) "Ranking Factors by Importance in Factorial Survey Analysis," *Communications of the Association for Information Systems*: Vol. 42 , Article 8.
DOI: 10.17705/1CAIS.04208
Available at: http://aisel.aisnet.org/cais/vol42/iss1/8

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Communications of the Association for Information Systems

Tutorial DOI: 10.17705/1CAIS.04208 ISSN: 1529-3181

Ranking Factors by Importance in Factorial Survey Analysis

Robert F. Otondo
Dept. of Management & Information Systems
Mississippi State University
bob.otondo@msstate.edu

Robert E. Crossler
Dept. of Management, Information Systems, & Entrepreneurship
Carson College of Business
Washington State University

Merrill Warkentin
Dept. of Management & Information Systems
Mississippi State University

Abstract:

Factorial survey analysis is a statistical technique with a long history of use in decision-oriented organizational and information systems (IS) research. The technique produces a collection of standardized regression coefficients that help one to rank survey factors by importance. However, such rankings may be invalid because a researcher might not account for two related issues: unequal factor (i.e., dimension) manipulation effect sizes and the inherent multilevel structure of factorial survey data. We address these concomitant issues by demonstrating the ranking problem in simulated datasets, explaining the ranking problem's underlying statistical causes, and justifying the use of remediating statistical methods. In particular, we focus on coding proportional to effect, a technique in which one consolidates corresponding dimension-level dummy (0, 1) variables into a single re-calibrated independent variable that is regressed on the dependent variable. One then uses the resulting standardized coefficients to rank the factors. We assess the advantages, disadvantages, and limitations of remediation techniques and offer suggestions for future information systems research.

Keywords: Factorial Survey Analysis, Scaling, Multilevel Modeling, Centering.

This manuscript underwent peer review. It was received 08/05/2016 and was with the authors for 12 months for 2 revisions. Joerg Evermann served as Associate Editor.

1 Introduction

The factorial survey method—also known as policy capturing, conjoint measurement, scenario studies, and experimental vignette methodology (Ongena, van de Wijngaert, & Huizer, 2013; Aguinis & Bradley, 2014)—assesses an individual's responses to decision processes. Factorial surveys combine controlled treatments from a multivariate experimental design with sampling techniques of survey procedures (Rossi & Anderson, 1982). Researchers construct treatments in a controlled design from discrete levels of salient orthogonal variables. Depending on the research question, researchers may construct treatments as hypothetical stories or objects (Rossi & Anderson, 1982). Treatments constructed as stories—also called *vignettes* or *conjoints*—often include fictional characters engaged in a specific behavior (Alexander & Becker, 1978). Treatments constructed as object descriptions—also called *profiles* (Krieger, Green, & Umesh 1998)—include salient characteristics and features (e.g., car profiles may list color, price, mileage, maximum speed, and length; Steckel, DeSarbo, & Mahajan, 1991). Subjects review one or more treatments and then complete a survey about each treatment. Analysis of factorial survey data describes statistical relationships between message cues and subject judgments, which one can use to rank the independent variables by importance.

Researchers have used factorial surveys across a variety of IS contexts, including IS policy compliance (D'Arcy, Hovav, & Galletta, 2009; Moores & Chang, 2006; Lee, Crossler, & Warkentin, 2013), communication media choice (Webster & Trevino, 1995), mobile technology adoption (Bouwman & van de Wijngaert, 2009), user-interface design testing (Vance, Lowry, & Eggett, 2015), and computer abuse and IS policy violations (Guo, Yuan, Archer, & Connelly, 2011; Vance, Lowry, & Eggett, 2013; Johnston, Warkentin, McBride, & Carter, 2016; Trinkle, Crossler, & Warkentin, 2014; Willison, Warkentin, & Johnston, 2016). Its popularity in information security research (Crossler et al, 2013) reflects its common use for studying deviant behaviors in criminology (Barlow, Warkentin, Ormond, & Dennis, 2013) and ethical issues in business decision making (O'Fallon & Butterfield, 2005). Researchers have also applied factorial survey analysis to other organizational contexts such as ethical decision making (Harrington, 1996; Trevino & Victor, 1992), professional reconciliation decisions (Tomlinson, Dineen, & Lewicki, 2004), job searches, compensation, employee discipline, job analysis, sexual harassment, employment interviews, and contract arbitration (Karren & Barringer, 2002).

Researchers often analyze factorial survey data using Rossi and Anderson's (1982) two-step method, which comprises a series of ordinary least squares (OLS) regressions. Though widely used, this two-step method has two major shortcomings. First, the method does not account for inadvertent unequal dimension manipulation effect sizes that result from the unequal dimension manipulation intervals. Such dissimilarities are important because they can create spurious results and potential confounds between artifacts of the research design and judgments of the sample's respondents. Second, the two-step method does not account for the inherent hierarchical, multilevel structure of factorial survey data. Instead, the two-step method—like other factorial survey methods—collapses multiple vignette judgments from each individual into a single collective data pool. Since each individual's own judgments of their assigned vignettes tend to be more similar vis-à-vis the aggregate judgments of all other individuals, collapsing multilevel data violates two important assumptions of OLS regression: 1) the independence of cases and 2) homoscedasticity of residuals (Hox, Kreft, & Hermkens, 1991). Collapsing also aggregates vignette-level and individual-level error variance into one residual, which means that within- and between-individual variation cannot be distinguished (Wallander, 2009). A second set of problems stems from individual variation in main effects (i.e., which one would assess based on variation in intercepts across individuals) and in rates of association or rates of change across vignettes (i.e., which one would assess based on variation in slopes across individuals). Thus, unmodeled multilevel structures can hide individual-level effects on vignette-level relationships that may change the values and meanings of results from analyses that do not use a multilevel approach.

Unfortunately, Rossi and Anderson's (1982) two-step method does not rectify these two problems. Consequently, spurious rankings can compromise theoretical generalizability and practical usefulness. Since practitioners may use the two-step method's results to guide organizational decision making (e.g., establishing fair price thresholds for childcare; Shlay, Tran, Weintraub, & Harmon, 2005; determining sanction certainty and/or severity in IS security practices; D'Arcy et al., 2009), it is important for researchers and practitioners to avoid spurious rankings that can produce adverse organizational outcomes.

In this paper, we address the above shortcomings and, thus, help to establish best practices for Rossi and Anderson's (1982) two-step method in particular and factorial survey analysis in general. We organize the paper into two parts. In the first part, we mathematically examine unequal dimension manipulation effect sizes, and, in the second part, we examine multilevel modeling techniques. In each part, we focus on the ranking of standardized coefficients for vignette dimensions and illustrating potential remediation. We also highlight applications and limitations of each technique in IS research.

2 Analyzing Factorial Survey Data

Factorial surveys are particularly useful when subjects may be reluctant to express their inclination to perform a deviant behavior. To mitigate potential social desirability bias, researchers instead ask subjects how they might respond if they were the vignette character or whether they would act in the same manner as the vignette character. When written in an indirect and unintimidating way, vignettes diffuse a subject's personal responsibility for performing a maladaptive act. In this way, factorial surveys can mitigate social desirability, acquiescence, or other cognitive and social biases while capturing a subject's true intentions. Appendix A displays an example of a full factorial survey vignette.

In step one of Rossi and Anderson's (1982) two-step method, a researcher regresses the dependent Y variable on dummy (0, 1) vignette-level independent variables that represent the absence or presence, respectively, of statements that describe a particular level for a particular dimension in a corresponding vignette. In the parlance of factorial survey analysis, a *dimension* refers to "a discrete variable associated with the phenomenon being studied" and a *level* is a "specific value within a dimension" (Shlay et al., 2005, p. 397). The vignette in Appendix A provides examples of dimensions and intervals. *Self-efficacy* is an example of a dimension. The vignette operationalizes levels in the self-efficacy dimension as follows: low self-efficacy by the statement "[Joe] is not confident he can do it again easily" and high self-efficacy by the statement "[Joe] is confident he can do it again easily". Each dimension level is, in turn, associated with a certain ΔY and a corresponding unstandardized regression coefficient β.

The researcher must then correct the unstandardized β coefficients from step one to account for unequal dimension-level intervals within each dimension. Thus, in step two, the researcher implements this correction through coding proportional to effect (CPE), a process that "recalibrate[s] the scale so the intervals are proportional to the 'effect' of each category [i.e., each dimension level]" (Boyle, 1970, p. 465). CPE involves replacing each dummy variable that represents a level included in the model (i.e., in which the dummy variable is equal to 1) with its corresponding unstandardized coefficient β from step one's regression. Levels that act as a reference group (i.e., those levels not included in the model or represented by a dummy value equal to 0) do not require that replacement because the corresponding β would equal zero. The researcher then sums the separate unstandardized coefficients for each dimension level from step one into a single independent variable that represents that one dimension and regresses the Y variable on these aggregated single independent variables. "This technique", explains Lauder (2002), "allows for an examination of the contribution of each dimension, relative to, and controlling for, all other dimensions in the model" (p. 40). The researcher can then use the standardized coefficients $\mathbf{B}$ of this step two regression as an index to rank vignette-level dimensions by their degree of influence on the Y variable. Note that, to do so, one must use the standardized coefficient $\mathbf{B}$—not the unstandardized β—because the latter will equal (or approximately equal) 1.00 in the step two regression.

3 Message Manipulations and Dimension Effect Sizes in Factorial Surveys

While the CPE technique accounts for unequal dimension-level intervals *within* each dimension, Rossi and Anderson's (1982) two-step method does not account for unequal intervals *between* dimensions (e.g., across factors being studied). A researcher may inadvertently introduce unequal dimension interval sizes—and corresponding dissimilarities in dimension manipulation effect sizes—when dimension-level statements are too limited, too extreme, or inconsistent across dimensions, which is particularly true of vignettes that incorporate meaningful but imprecise qualitative descriptors. Indeed, creating message content and cues can be challenging because the researcher must balance the needs of 1) producing a vignette storyline that is reasonably believable by the sample pool and 2) manipulating subjects sufficiently to produce acceptable variation in the dependent variable. Since the wording of each dimension-level statement is a product of the individual research designer, confounds between treatment

and subject effects can occur (Netemeyer, Bearden, & Sharma, 2003). As we show next, significant dissimilarities in dimension manipulations levels can compromise the validity of dimension ranking in CPE.

One way to demonstrate how differing manipulation effect sizes arise from unequal dimension intervals is to examine the mathematical relationship between standardized and unstandardized regression coefficients. Standardized regression coefficients are calculated as follows:

$$\mathbf{B}_x = \frac{\beta_x * s_x}{s_Y}, \quad (1)$$

where $\mathbf{B}_x$ is the standardized regression coefficient of x, β_x the unstandardized coefficient of x, s_x the standard deviation of x, and s_y the standard deviation of y.

Equation 1 indicates that both β_x and s_x must be assessed to calculate $\mathbf{B}_x$. As we discuss above, the two-step method addresses differences in *within-dimension* manipulation effect sizes (i.e., the differences in ΔY intervals across dimension-level statements) by substituting step one's β_x values for dummy-level values in the CPE calculation of $\mathbf{B}_x$ in step two. Unfortunately, neither step one nor step two account for variations in dimension interval size (i.e., as s_x in Equation 1 represents), and, therefore, neither addresses differences in *between-dimension* manipulation effect sizes (i.e., the differences in ΔY intervals between the factors being studied). This shortcoming means that *within-dimension* manipulation effects are equalized but *between-dimension* manipulation effect sizes are not. We explore the ramifications of unequal dimension intervals—and their corresponding effects on dimension manipulation effect size and dimension ranking—in our examples that use Datasets 1-3.

Note that a researcher can mitigate the impact of unequal dimension intervals in several ways, such as grounding message cues on current practice, economic costs, or other organizational factors to create realistic vignette storylines (Dallal, 2001). Furthermore, the researcher should perform such mitigation efforts during the research design phase of the study because CPE and multilevel analysis cannot account for unequal effect sizes between dimensions. Thus, Dallal's (2001) suggestions are similar to many common research-design practices, such as keeping surveys short to minimize survey fatigue and non-response bias (Kriauciunas, Parmigiani, & Rivera-Santos, 2011) and obtaining measures from different sources to minimize common method variance (Podsakoff, MacKenzie, & Lee, 2003).

We demonstrate the consequences of unequal dimension manipulation effect sizes on dimension ranking in CPE to underscore the importance of such mitigation efforts. We created three different single-level (i.e., non-multilevel) datasets, all inspired by Johnston et al.'s (2016) example in Appendix A. Briefly, that example—which draws on deterrence theory in social psychology and criminology—has shown that perceptions of the certainty, severity, and celerity (i.e., swiftness) of punishment are positively associated with intention to comply with information security policies. That is, individuals are more likely to follow organizational policies when they see punishments as severe, certain, and/or swift. Johnston et al. (2016), like other researchers who have conducted research in criminology and information security (e.g., Herzog, 2003; Vance, Lowry, & Eggett, 2013), investigated those relationships using factorial survey analysis. Accordingly, we use perceptions of *Certainty*, *Severity*, and *Celerity* as dimensions (i.e., independent variables) and *Intention to Comply* (*IntComply*) with information security policies as the dependent variable.

In the two-step method, dummy variables set at 0 or 1 represent the absence or presence, respectively, of dimension-interval statements in the vignette. Beta coefficients from the regressions in the two-step method indicate the effect of a manipulation dummy variable on the outcome variable *IntComply*. Importantly, in our demonstration, we assume that our qualitative variables are based on underlying quantitative scales that our hypothetical researcher does not know about. Thus, we ask the reader to keep this important assumption in mind as we discuss the quantitative aspects of our demonstration.

Table 1 summarizes the parameter specifications for Datasets 1, 2, and 3. Datasets 1 and 2 demonstrate two different experiments that have equivalent dimension intervals (and dimension manipulation effect sizes) for *Celerity* that are measured in two alternative but interconvertible scales (e.g., when one study mentions timeframes in days but another in weeks). In Table 1, the dimensions levels of low, medium, and high *Celerity* (i.e., *Cel_Low*, *Cel_Med*, and *Cel_Hi* in the third row) in Dataset 1 are equal to one-third of those in Dataset 2, while the β_1 value for Dataset 1 (i.e., the *Celerity* coefficient in the second row in Table

1) is three times that of Dataset 2. Thus, the differences balance out; for example, the partial effect of β_1 * *Cel_Med* in Dataset 1 equals 1.50 * 0.50 = 0.75, while the partial effect of β_1 * *Cel_Med* in Dataset 2 equals 0.50 * 1.50 = 0.75. Note that we set the differences between low-med and med-high intervals for each dimension in Table 1 to be equal for ease of calculation and presentation. We would not necessarily expect such equalities in real-world data.

Table 1. Parameters Used to Generate Datasets 1, 2, and 3

Vignette-level parameter	Dataset 1	Dataset 2	Dataset 3
IntComply-intercept (β_0)	0.00	0.00	0.00
β_1 (*Celerity* coefficient)	1.50	0.50	1.50
Cel_Low / *Cel_Med* / *Cel_Hi* (dimension-level) effects	0.00 / 0.50 / 1.00	0.00 / 1.50 / 3.00	0.00 / 1.50 / 3.00
β_2 (*Certainty* coefficient)	1.25	1.25	1.25
Crt_Low / *Crt_Med* / *Crt_Hi* (dimension-level) effect size	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00
β_3 (*Severity* coefficient)	2.00	2.00	2.00
Sev_Low / *Sev_Med* / *Sev_Hi* (dimension-level) effect size	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00
Residual (e_{ti})	0.00 (0.40)	0.00 (0.40)	0.00 (0.40)

Notes: we include standard deviations (in parentheses) where applicable (e.g., for residuals). *N* = 10,800 (400 individuals, 27 vignettes each). Coefficients and dimension-level effect sizes for variables *Certainty* and *Severity* (e.g., β_2, *Crt_Low*, and *Crt_Hi*; β_3, *Sev_Low*, and *Sev_Hi*) remain constant across all three datasets.

Datasets 1 and 3, on the other hand, demonstrate unequal dimension interval effect sizes. The dimension levels for *Cel_Low*, *Cel_Med*, and *Cel_Hi* (third row) in Dataset 3 are three times those of Dataset 1, but the relationship between *Celerity* and *IntComply* (i.e., the respective *Celerity* coefficients, β_1, in the second row) are equal. Thus, the differences do not balance; for example, the partial effect of β_1 * *Cel_Med* in Dataset 1 equals 1.50 * 0.50 = 0.75, while the partial effect of β_1 * *Cel_Med* in Dataset 3 equals 1.50 * 1.50 = 2.25. Dimension-level intervals, manipulation effects, and coefficients for the low-, medium-, and high-level dummy variables for *Certainty* and *Severity* (i.e., *Crt_Low*, *Crt_Med*, *Crt_Hi*, *Sev_Low*, *Sev_Med*, and *Sev_Hi*, respectively) are constant across all three datasets.

We generated the three datasets using R version 3.3.2. We generated them across 400 individuals with each individual associated with all 27 vignettes from our 3 x 3 x 3 design. This research design produces a vignette-level sample size of 10,800. Figure B1 in Appendix B provides the code we used to develop these datasets, and Tables B1 and B2 provide the descriptive statistics.

Given that researchers commonly use Rossi and Anderson's (1982) two-step method to analyze factorial survey data, we begin our study with OLS regressions—which that method uses—in order to establish a baseline for comparing current practice with multilevel modeling in Section 4. Tables 2 and 3 display the results from the OLS analyses of Datasets 1, 2, and 3. We label the models as follows. The first number refers to the table number (i.e., Table 2 or 3); Tables 2 and 3 display results from steps one and two, respectively, of Rossi and Anderson's (1982) two-step method. The second number represents the dataset the model used. We used the low-level variables of each dimension (i.e., *Cel_Low*, *Crt_Low*, and *Sev_Low*) as the reference levels for these analyses; therefore, we omitted them to avoid linear dependencies. We used IBM SPSS Statistics version 24 to perform the analyses. Table 2 (step one) shows that equivalent dimension manipulation effect sizes produced from interconvertible scales—with coefficients adjusted accordingly—in Datasets 1 and 2 produce similar coefficient estimates for the *Celerity* dummy variables (i.e., the *Cel_Hi* coefficient, β_1, ≈ 1.50 and the *Cel_Med* coefficient, β_2, ≈ 0.75 in Models 2-1 and 2-2). In turn, variables *Celerity*, *Certainty*, and *Severity* are ranked in the same order in Models 3-1 and 3-2 in step two in Table 3 (i.e., *Severity* > *Celerity* > *Certainty*).

Table 2. Comparison of Step One Analyses of Datasets 1-3

Dummy variables (coefficients)	Unstandardized coefficients (β)					
	Model 2-1 (Dataset 1)		Model 2-2 (Dataset 2)		Model 2-3 (Dataset 3)	
	True value[a]	Simulated[b]	True value[a]	Simulated[b]	True value[a]	Simulated[b]
IntComply-intercept (β_0)	0.000	-0.011 (0.010)	0.000	0.006 (0.010)	0.000	-0.008*** (0.010)
Cel_Hi (β_1)	1.500 [1.50 x 1.00]	1.506*** (0.010)	1.500 [0.50 x 3.00]	1.505*** (0.009)	4.500 [1.50 x 3.00]	4.509*** (0.0109)
Cel_Med (β_2)	0.750 [1.50 x 0.50]	(0.747*** 0.010)	0.750 [0.50 x 1.50]	0.746*** (0.009)	2.250 [1.50 x 1.50]	2.265*** (0.009)
Crt_Hi (β_3)	1.250 [1.25 x 1.00]	1.249*** (0.010)	1.250 [1.25 x 1.00]	1.245*** (0.009)	1.250 [1.25 x 1.00]	1.254*** (0.009)
Crt_Med (β_4)	0.625 [1.25 x 0.50]	0.627*** (0.010)	0.625 [1.25 x 0.50]	0.621*** (0.009)	0.625 [1.25 x 0.50]	0.631*** (0.009)
Sev_Hi (β_5)	2.000 [2.00 x 1.00]	2.012*** (0.010)	2.000 [2.00 x 1.00]	1.997*** (0.009)	2.000 [2.00 x 1.00]	1.991*** (0.009)
Sev_Med (β_6)	1.000 [2.00 x 0.50]	1.003*** (0.010)	1.000 [2.00 x 0.50]	0.994*** (0.009)	1.000 [2.00 x 0.50]	0.987*** (0.009)

Notes: N = 10,800 at vignette level; N = 400 at individual level.
[a] We show calculations for true values (i.e., dimension coefficients for *Celerity*, *Certainty*, or *Severity* multiplied by dimension-level manipulation effect *hi* or *med*) in square brackets.
[b] We show unstandardized coefficient estimates, standardized coefficient estimates, and robust standard errors for simulated coefficients in round brackets.
*** $p < .001$, ** $p < .01$, * $p < .05$.

Table 3. Comparison of Step Two (CPE) Analyses of Datasets 1-3

Aggregate dimension variables (coefficients)	Model 3-1 (Dataset 1)		Model 3-2 (Dataset 2)		Model 3-3 (Dataset 3)	
	Unstd. (β) coefficient	Stand. (**B**) coefficient	Unstd. (β) coefficient	Stand. (**B**) coefficient	Unstd. (β) coefficient	Stand. (**B**) coefficient
IntComply-intercept (β_0)	-0.011 (0.009)		0.006 (0.009)		-0.008 (0.009)	
Celerity (β_1)	1.000*** (0.006)	0.506***	1.000*** (0.006)	0.508***	1.000*** (0.002)	0.870***
Certainty (β_2)	1.000*** (0.008)	0.420***	1.000*** (0.008)	0.420***	1.000*** (0.008)	0.242***
Severity (β_3)	1.000*** (0.005)	0.676***	1.000*** (0.005)	0.674***	1.000*** (0.005)	0.384***

Notes: N = 10,800 at vignette level; N = 400 at individual level.
[a] We show calculations for true values (i.e., dimension coefficients for *Celerity*, *Certainty*, or *Severity* multiplied by dimension-level manipulation effect *hi* or *med*) in square brackets.
[b] We show unstandardized coefficient estimates, standardized coefficient estimates, and robust standard errors for simulated coefficients in round brackets.
*** $p < .001$, ** $p < .01$, * $p < .05$.

Comparing the step one results in Models 2-1 and 2-3 (i.e., Datasets 1 and 3, respectively) reveals different β_1 coefficient estimates for their *Cel_Hi* dummy variables and different β_2 coefficient estimates for their *Cel_Med* dummy variables. These differences arise because the dimension interval effect sizes in Dataset 3 are three times those of Dataset 1 even though both datasets are based on the same "true value" for the *Celerity* coefficient (i.e., 1.50; Table 1, second row). Figure 1 illustrates how these unequal dimension manipulation intervals for the *Celerity* variable can generate such differences.

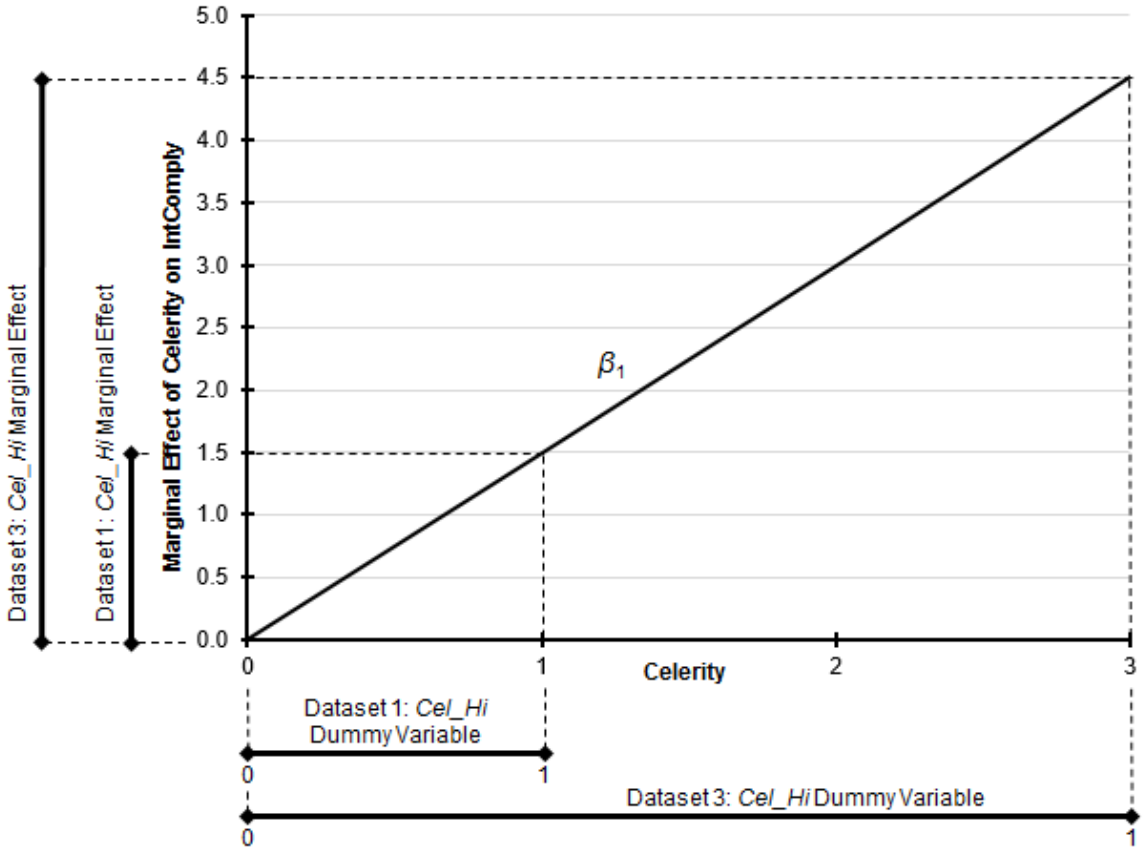

Figure 1. Comparative Marginal Effects of Cel_Hi Dummy Variable on Intention to Comply, Dataset 1 vs. 3

Figure 1 draws on Table 1 (i.e., β_1 in Table 1, Row 2) to illustrate the regression of *IntComply* on the underlying *Celerity* factor after controlling for *Certainty* and *Severity*. In other words, the vertical axis shows the marginal effect of *Celerity* on *IntComply*. Figure 1 also displays the marginal effects of the *Cel_Hi* dummy variable intervals from Datasets 1 and 3 on *IntComply*: to the left of the graph in Figure 1 are lines that represent the marginal values of *IntComply* from Datasets 1 and 3; to the bottom are lines that represent the relationship between the *Celerity* scale and the *Cel_Hi* dummy variables from Datasets 1 and 3. An important characteristic of the *Cel_Hi* dummy variable lines is that, while they represent different *ranges of values* on the *Celerity* scale, *the same dummy variable value* (i.e., 1) represents them in the step one regression equations. In other words, a "1" value that signifies the *Cel_Hi* dummy variable in Dataset 3 represents three times the effect of *Celerity* than the "1" value that signifies the *Cel_Hi* dummy variable in Dataset 1. Consequently, there is also a three-fold increase in the *Cel_Hi* regression coefficient in Dataset 3 over that of Dataset 1 (i.e., β_1, Table 2, Row 2). This three-fold difference in the β_1 coefficients will eventually—and adversely—affect the rankings of the standardized **B** coefficients in step two. Thus, Figure 1 demonstrates how unequal dimension intervals create inconsistent rankings of the standardized **B** coefficients in Rossi and Anderson's (1982) two-step method.

Mathematically comparing the sequential or Type I sum of squares for *Cel_Hi* and *Cel_Med* explains these effects (Appendix C). This comparison has at least two important ramifications. First, comparing the models in Tables 2 and 3 shows that larger dimension intervals inflate both the unstandardized β and standardized **B** coefficients of *Celerity* in steps one and two, respectively. In turn, the inflated standardized **B** coefficient, if used to rank coefficients, would place *Celerity* higher than *Certainty* and *Severity* in Model 3-3 but lower than *Severity* in Model 3-1. Again, the difference in ranking would arise due to the unequal dimension interval and manipulation effect sizes of *Celerity* rather than per unit association between *Celerity* and the dependent variable *IntComply*. The situation is of little concern when a researcher uses transformable interval scales (e.g., Model 3-1 vs. 3-2) because the β coefficients will account for such differences (e.g., the coefficient for weeks will be seven times that of the coefficient for days).

The second ramification is that unequal dimension interval and manipulation effect sizes may be more likely in qualitative statements when wording or meaning is inconsistent or ambiguous across dimensions. This latter conclusion strongly questions the generalizability of rankings across studies. A more reliable way to mitigate this problem involves replacing qualitative statements and their dummy (0, 1) variables with more precise statements that contain quantified dimension levels as the policy-capturing technique does (e.g., Stewart, 1988; Aiman-Smith, Scullen, & Barr, 2002; Karren & Barringer, 2002). The policy-capturing design avoids the CPE process altogether.

A third ramification is that the validity of multilevel modeling of factorial surveys will rely ultimately on the underlying research design. Researchers must, therefore, take care to ensure the quality of the dimensions and levels used in the vignettes and judiciously consider the impact of unequal dimension intervals and manipulation effects on CPE rankings—especially if the **B** values are close. Researchers should also acknowledge the possibility that unequal dimension manipulations are a limitation of their research.

The examples we present in Tables 1-3 suggest that quantified dimensions (i.e., as found in the policy-capturing method) are superior to CPE-based standardized coefficients as a means for ranking vignette dimensions. Our analyses are consistent with previous research; for example, Dallal (2001) argues that comparisons of standard coefficients to determine importance can be "illusory" when rankings depend on "the specific context and why the question is being asked" or how easily the subject can modify the value of an independent variable (e.g., "time watching television" versus "weight or cholesterol level" and "height or age"). Dallal (2001) opines that: "This is not an issue if the question asks what most determines the response, but it is critical if the point of the exercise is to develop a public policy to effect a change in the response". Dallal's argument is also germane to organizational contexts. For example, improving information security can involve managerial costs of policy and procedure updates, hardware and software investments for enhancing detection certainty, personnel costs from shortening process times and improving celerity, and administrative costs for adjudicating increased appeals due to more severe punishments.

Though quantitative variables are more desirable, they may not be feasible in all situations. In such cases, the researcher should pattern treatment dimension levels after actual organizational practices. Such advice means extra work for the research team, but it is worthwhile for at least two reasons. First, the resulting experiment will align more faithfully to organizational reality. Second, more realistic experiments help set the stage for increasing the validity of a study's contributions to theory and practice (e.g., experiments that the dimension manipulation effect size of low-ranking factors may lead to recommendations on how to improve organizational outcomes).

4 Multilevel Modeling of Factorial Survey Data

As we state above, two of the most egregious problems of failing to use multilevel modeling are violations of the assumptions of independence of cases and homoscedasticity of residuals. Multilevel analysis avoids these problems in factorial survey analyses by modeling individual-level characteristics and their cross-level effects on vignette-level characteristics and responses. Indeed, past IS research has demonstrated the value of multilevel modeling (e.g., Bélanger & Crossler, 2011; Burton-Jones & Gallivan, 2007; Mithas, Ramasubbu, Krishnan, & Fornell, 2006-7), including some research that has used factorial survey methods (e.g., Vance et al., 2015; Bouwman & van de Wijngaert, 2009). However, these latter multilevel modeling efforts have largely only controlled for the effects of age, gender, and other demographic variables, which leaves much fertile ground for future factorial survey-based IS research. IS researchers could also use individual-level factors such as attitudes and personality traits (e.g., Hurt et al., 1999) and self-efficacy and fear of failure (e.g., Mitchell & Shepherd, 2010), which multilevel management and entrepreneurship researchers have used, in their research. We hope the following sections inspire such efforts.

Note that multilevel modeling does not mitigate the effects of the unequal dimension effect sizes (see Section 3). Indeed, as we say above, researchers can mitigate those unwanted effects by grounding message cues on current practice, economic costs, or other organizational factors to create realistic vignette storylines (Dallal, 2001). However, they must take these measures during research design (i.e., *before* collecting data) because they cannot account for them with multilevel modeling and methods.

4.1 Construction of Simulated Dataset 4

We demonstrate the dangers of ignoring multilevel structures in factorial surveys—and their consequent effects on CPE—by extending Hox et al.'s (1991) example to include cross-level moderation effects. We demonstrate factorial survey analysis and CPE using random coefficients modeling (RCM), a regression-based type of multilevel modeling. Our models are based on three level-one dimensions from the example in Appendix A (i.e., *Celerity*, *Certainty*, and *Severity*). We represent each dimension with three levels (e.g., *Cel_Low*, *Cel_Med*, and *Cel_Hi*) and model each one with their corresponding dummy (0, 1) variables. Our models also incorporate a continuous individual-level (level-two) variable, *risk aversion (RiskAverse)*, which the example in Appendix A suggests (i.e., "persons exhibiting a strong stability meta-trait are found to be more risk-averse and may avoid actions that place them at risk for threat and sanction-related consequences"; Johnston et al., 2016, p. 243). In our simulated datasets, *RiskAverse* exerts a direct cross-level effect on the level-one intercept and a positive moderating effect on the relationships between the *Celerity* dimension levels and the dependent variable *IntComply*. That is, individuals with high levels of *RiskAverse* will also exhibit strong positive associations between *Celerity* and *IntComply*, while individuals with low levels of *RiskAverse* will also exhibit weak positive associations between *Celerity* and *IntComply*. Further, π_q and β_{qs} represent vignette-level (level-one) and individual-level (level-two) coefficients, respectively: q represents the predictor's level-one order (0 to Q), and s represents the predictor's level-two order (0 to S). Note that while q applies to both levels, s applies only to level-two predictors.

We generated Dataset 4 using R version 3.3.2. We generated the dataset across 400 individuals with each individual associated with all 27 vignettes. This design produced a vignette-level sample size of 10,800. The vignette-level (level-one) magnitudes and coefficients in Dataset 4 mirror those in Dataset 1. Figure D1 in Appendix D provides the code we used to generate Dataset 4.

4.2 An OLS Modeling of Dataset 4

We analyzed Dataset 4 using the traditional two-step OLS regression method of factorial survey analysis to provide a baseline comparison for our multilevel analyses. The low-level variables of each dimension (i.e., *Cel_Low*, *Crt_Low*, and *Sev_Low*) served as reference levels for this analysis; as such, we deleted them from subsequent models to avoid linear dependencies. We used IBM SPSS Statistics version 24 for the analyses. Tables 4 and 5 display the results of the step one and step two analyses. In step one, one regresses the dependent variable *IntComply* on dummy (0, 1) vignette-level independent variables that represent the absence or presence, respectively, of statements that describe a particular level for a particular dimension in a corresponding vignette. In step two, one accounts for unequal dimension-level intervals within each dimension via CPE. Briefly, the CPE process replaces each level that has a dummy variable equal to 1 with its corresponding unstandardized coefficient β from the step one regression and then sums each dimension's βs into a single independent variable. Levels that have a dummy variable equal to zero act as reference groups; as such, the step one regression does not include them, and they are not relevant to step two. The researcher then regresses the recalibrated dimension scales on the independent variable and uses the resulting standardized $\mathbf{B}_x$ coefficients as an index to rank dimensions by importance. The standardized $\mathbf{B}_x$ coefficients in Table 5 indicate the most important factor is *Celerity* followed in order by *Severity* and *Certainty*. Note that this ranking differs from the coefficients we used to construct Dataset 4 (i.e., *Celerity*: β_1 = 1.50, *Certainty*: β_2 = 1.25, and *Severity*: β_3 = 2.00; Table D2, Appendix D). This difference is not unexpected given the positive moderating effects of *RiskAverse* on the relationship between *Celerity* and the dependent variable *IntComply*. While one could argue that the results in Table 5 do correctly rank *Celerity* as the most important coefficient, one can see that OLS regression obscures *RiskAverse*'s effect on the *Celerity* coefficient. We explain this relationship—and its consequences on CPE ranking—in more detail in Section 4.3.

Table 4. OLS Estimates from Multilevel Factorial Survey Data (Step One, Dataset 4)

Variables (coefficients)	True values (from Table 1)	Unstandardized coefficients (β)
IntComply-intercept (β_0)	0.000	0.702*** (.019)
Cel_Hi (β_1)	1.500	2.677*** (.018)
Cel_Med (β_2)	0.750	1.324*** (.018)
Crt_Hi (β_3)	1.250	1.259*** (.018)
Crt_Med (β_4)	0.625	0.646*** (.018)
Sev_Hi (β_5)	2.000	2.008*** (.018)
Sev_Med (β_6)	1.000	1.003*** (.018)

Notes: N = 10,800 at vignette level; N = 400 at individual level. We show unstandardized coefficient estimates and robust standard errors in parentheses. Adjusted R^2 = 0.79.
*** $p < .001$, ** $p < .01$, * $p < .05$

Table 5. OLS Estimates from CPE Multilevel Factorial Survey Analysis (Step Two, Dataset 4)

Dimension	Regression coefficients			
	Unstandardized (β)	Std. Error	Standardized (β_x)	Rank
IntComply-intercept (β_0)	0.702***	0.017		
Celerity	1.000***	0.007	0.665	1
Certainty	1.000***	0.014	0.313	3
Severity	1.000***	0.009	0.499	2

Notes: N = 10,800 (400 individuals, 27 vignettes each). Adjusted R^2 = 0.79. Values for the independent variables are unstandardized regression coefficients from Table 4. Rank of importance based on standardized coefficients.
*** $p < .001$, ** $p < .01$, * $p < .05$.

4.3 Modeling Dummy (0, 1) Variables in Multilevel Factorial Survey Analysis

Our RCM approach follows the traditional procedure of constructing the unconditional model first followed by a series of conditional models that test the effects of level-one and level-two variables (Raudenbush & Bryk, 2002). We constructed and analyzed all random coefficients models with HLM for Windows version 7.01 using full maximum likelihood estimation. As in our unequal dimension manipulation effect sizes examples in Section 2, we assume that our hypothetical researcher does not know the underlying quantitative scales used to construct Dataset 4. Again, we ask the reader to keep this assumption in mind as we discuss our next examples.

Our RCM analysis of Dataset 1 begins with step one of the two-step factorial survey analysis procedure. We use four random coefficient models, all of which use uncentered level-one variables. The 0.00 values for *Celerity*, *Certainty*, and *Severity* represent the *Cel_Low*, *Crt_Low*, and *Sev_Low* dimension levels, respectively. The first model (i.e., Model 6-1, Table 6) is a fully unconditional model that contains no level-one or level-two predictors. Thus, it serves as a baseline from which improvements in subsequent models can be assessed (Raudenbush & Bryk, 2002). Model 6-2 adds the six "*Hi*" and "*Med*" level-one dummy variables (e.g., *Cel_Hi*, *Cel_Med*). We omit the "*Low*" dimension levels (i.e., *Cel_Low*, *Crt_Low*, and *Sev_Low*) so they can serve as reference points for their corresponding dummy (0, 1) variables. Model 6-3 assesses the direct and moderating effects of the (uncentered) level-two predictor *RiskAverse*. *RiskAverse*'s direct effect is assessed on π_0 (i.e., the *IntComply*-intercept), while its moderating effects are assessed on the six dummy (0, 1) coefficients (i.e., π_1 to π_6). Figure E1 in Appendix E provides an

example of the multilevel equations in the HLM interface for Model 6-4 that can help explain the values listed in Table 6.

Table 6. RCM Analysis of Dummy (0, 1) Variables from Factorial Survey: Step One (Dataset 4)

Variables (coefficients)	True value[a]	Model 6-1	Model 6-2	Model 6-3	Model 6-4
Level-one parameters $\pi_0 - \pi_6$ (corresponding level-two intercept $\beta_{00} - \beta_{60}$)					
π_0 (β_{00}, level-two intercept for *IntComply* intercept π_0)	0.00	3.674*** (.030)	0.702*** (.021)	-0.035 (.035)	-0.024 (.030)
π_1 (β_{10}, level-two intercept for *Cel_Hi* coefficient π_1)	1.50 [1.50 x 1.00]		2.677*** (.029)	1.497*** (.039)	1.497*** (.039)
π_2 (β_{20}, level-two intercept for *Cel_Med* coefficient π_2)	0.75 [1.50 x 0.50]		1.324*** (.016)	0.753*** (.027)	0.753*** (.027)
π_3 (β_{30}, level-two intercept for *Crt_Hi* coefficient π_3)	1.25 [1.25 x 1.00]		1.259*** (.009)	1.270*** (.023)	1.259*** (.009)
π_4 (β_{40}, level-two intercept for *Crt_Med* coefficient π_4)	0.625 [1.25 x 0.50]		0.646*** (.009)	0.650*** (.024)	0.646*** (.009)
π_5 (β_{50}, level-two intercept for *Sev_Hi* coefficient π_5)	2.00 [2.00 x 1.00]		2.008*** (.009)	2.014*** (.022)	2.008*** (.009)
π_6 (β_{60}, level-two intercept for *Sev_Med* coefficient π_6)	1.00 [2.00 x 0.50]		1.003*** (.009)	1.017*** (.025)	1.003*** (.009)
Direct and moderating effects of uncentered *RiskAverse* on level-one coefficients $\pi_0 - \pi_6$					
β_{01} (π_0, *IntComply*-intercept)	0.25			0.252*** (.035)	0.248*** (.009)
β_{11} (π_1, *Cel_Hi*)	0.40 [0.40 x 1.00]			0.403*** (.011)	0.403*** (.012)
β_{21} (π_2, *Cel_Med*)	0.20 [0.40 x 0.50]			0.195*** (.009)	0.195*** (.009)
β_{31} (π_3, *Crt_Hi*)	0.00			-0.004 (.007)	
β_{41} (π_4, *Crt_Med*)	0.00			-0.001 (.008)	
β_{51} (π_5, *Sev_Hi*)	0.00			-0.002 (.007)	
β_{61} (π_6, *Sev_Med*)	0.00			-0.005 (.008)	
Goodness of fit					
Deviance	NA	40,780.3	15,507.9	11,946.5	11,953.7
Number of parameters	NA	3	9	43	17
Δ Deviance from previous model	NA		-25,272.4***	-3,561.4***	N/A
Δ Deviance from Model 6-2	NA			-3,561.4***	-3,554.2***

Note: N = 10,800 at vignette level; N = 400 at individual level. We show unstandardized coefficient estimates and robust standard errors in parentheses. Level-one and level-two variables are uncentered (Hox et al. 1991).
[a] True values are the values used to generate Dataset 4, which is based on Dataset 1. We show calculations for true values (i.e., dimension coefficients for *Celerity*, *Certainty*, or *Severity* multiplied by dimension-level manipulation effect *hi* or *med*) in square brackets.
*** $p < .001$, ** $p < .01$, * $p < .05$.

Model 6-4 is our "final model" (Singer & Willett, 2003) that assesses potential changes in Model 6-3 when non-significant predictors are removed.

Results in Table 6 demonstrate the value of multilevel modeling. Model 6-2, which does not include the level-two predictor *RiskAverse*, overestimates the coefficients for *Cel_Hi* and *Cel_Med* (i.e., 2.677 and 1.324, respectively; standard errors = 0.029 and 0.016, respectively). That is, the values used to generate Dataset 4 (Table D2, Appendix D) are not within the 95 percent confidence intervals (CIs) estimated in Model 6-2. As expected, the coefficient estimates in Model 6-2 match those of the OLS regression model

in Table 4. However, when *RiskAverse* is added as an uncentered level-two predictor in Models 6-3 and 6-4, the 95 percent CIs of the level-1 *Cel_Hi* and *Cel_Med* coefficients and the *IntComply*-intercept (i.e., β_{00}) contain the respective inputs used to generate Dataset 4 (i.e., Table D2, Appendix D).

The increased accuracy misleadingly appears to come at the cost of increased level-one standard errors for the fixed parameters (i.e., the intercept and the *Certainty* and *Severity* coefficients). In fact, multilevel models such as Model 6-3 *correctly* estimate level-one standard errors because they partition individual- and group-level variance (Maas & Hox, 2004). Moreover, there are at least two adverse ramifications of modeling multilevel data with single-level methods. First, it creates a "well-known" downward bias of level-one standard errors (Maas & Hox, 2004, p. 128). Second, it underestimates variance and standard errors, which can then elevate the chance of Type I errors—sometimes substantially (Kreft, 1994). In short, the failure to use multilevel modeling means the researcher may "find effects where they don't exist" (Kelloway, n.d., p. 36). Though the increases in many level-one standard errors are three to four times larger in Model 6-3 than in Model 6-2 and relatively small compared to the sizes of the β_{qs} coefficients, these comparative sizes of level-one standard errors and β_{qs} coefficients are artifacts of our simulated data that we constructed for ease of demonstration. These comparative sizes should *not* be considered a general rule.

The results in Table 6 indicate that multilevel modeling will affect the ranking of *Celerity*, *Certainty*, and *Severity* by importance in CPE because the CPE calculations used *RiskAverse*'s effects, as measured by the coefficients $\beta_{01} - \beta_{61}$, as inputs. In particular, the higher π_1 and π_2 values in Model 6-2 (i.e., the single-level model) suggest that *Celerity* will be ranked *higher* than *Severity*, whereas the lower π_1 and π_2 values in Models 6-3 and 6-4 (i.e., the multilevel models) suggest that *Celerity* will be ranked *lower* than *Severity*.

Table 7 (next page) displays the CPE (step two) results, which are based on the unstandardized π coefficients from Model 6-4. We calculated the values in Table 7 using HLM for Windows version 7.01 using full maximum likelihood estimation. HLM for Windows does not provide standardized **B** values, so we used the standard deviation values $s_{Celerity}$, $s_{Certainty}$, $s_{Severity}$, and $s_{IntComply}$—which the HLM for Windows' multivariate data matrix (MDM) file provides—to calculate standardized **B** values via Equation 1. As expected, the standardized **B** values for *Certainty* and *Severity* in Table 7 remained largely the same as those from the step two OLS regression in Table 5. However, the standardized **B** values for *Celerity* differed substantially (i.e., 0.665 vs. 0.372 in Tables 5 and 7, respectively).

The standardized **B** values in Table 7 show that multilevel modeling accounts for the cross-level effects of *RiskAverse* on π_0 and π_1. That is, the order of the **B** values corresponds to that of the true values used to construct Dataset 4. These results suggest that researchers should test for potential multilevel effects to help ensure correct ranking of the level-one dimensions. In turn, such rankings not only provide evidence of the relative importance of the level-one variables with respect to the dependent variable but also help illuminate the influence of individual-level factors on vignette-level relationships between design factors and judgmental outcomes. Such results can provide rich, nuanced insights into theory and practice[1].

4.4 Centering Effects in Multilevel Factorial Survey Analysis and CPE

A related but often unappreciated decision in multilevel modeling is the centering of variables. While addressed in much organizational research (e.g., Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995), centering has not received wide attention in factorial survey analyses. Centering decisions—especially for level-two variables—are important because they affect the values and meanings of level-one coefficients, which can, in turn, affect the ranking of their respective variables.

Following recommendations in Hox et al. (1991), we have *not* centered the level-one dummy (0, 1) vignette variables in our previous models. Their advice is consistent with other recommendations to avoid centering categorical variables (e.g., Chen, 2014). Thus, our approach differs from common centering practices in RCM where level-one and level-two predictors are typically grand-mean centered (GMC) or centered within clusters (CWC). Grand-mean centering an independent variable "yields an intercept equal to the expected value of Y_{ij} for a person with a score on X equal to the mean across all individuals in the sample" (Hofmann & Gavin, 1998, p. 627). Grand-mean centering is perhaps the more widely recommended of the two. It provides a more easily interpreted estimate of the Y-intercept, especially

[1] We thank an anonymous reviewer for this suggestion, which we paraphrase here.

when X_{ij} = 0 is either outside the sample range or is an unrealistic or spurious value (e.g., immediate punishment of an ethical violation). Grand-mean centered variables also offer some computational advantages over uncentered forms (Hofmann & Gavin, 1998). On the other hand, centering within clusters—also called group-mean centering—of level-one variables is often preferred over grand-mean centering because the latter confounds a portion of the within- and between-cluster variation. CWC avoids this confound by centering the level-one variables on the mean of their corresponding level-two cluster (e.g., individuals in factorial survey designs). The CWC process "removes all between-cluster variation from the predictor and yields a 'pure' estimate of the pooled within-cluster (i.e., Level-1) regression coefficient" (Enders & Tofighi, 2007, p. 128).

Using uncentered level-one dummy variables in factorial survey analysis has several advantages. First, it clearly identifies which dimension levels a vignette includes and which it does not, which helps one interpret the model and its results. Second, using uncentered level-one dummy variables allows the level-one intercept π_0[2] to represent the expected Y value from a baseline vignette in which all dummy (0, 1) values equal zero. Nonetheless, centering level-one dummy variables may be useful in some situations. We encourage interested readers to read Enders and Tofighi (2007, pp. 134-135, 138) for an in-depth discussion on the topic.

Table 7. RCM Analysis of Dummy (0, 1) Variables from Factorial Survey: Step Two (Dataset 4)[a]

Parameters		CPE process		
		Step one inputs[b]	Step two inputs[b]	
Coefficient	**True value[c]**	**From Model 6-4 β**	**β**	**B[d]**
π_0 (β_{00}, level-two intercept for *IntComply* intercept π_0)	0.00	N/A	-0.024 (0.029)	
π_1 (β_{10}, level-two intercept for *Celerity* coefficient π_1)	*Cel_Hi*: 1.50 *Cel_Med*: 0.75	*Cel_Hi*: 1.497 *Cel_Med*: 0.753	1.000*** (0.026)	0.372***
π_2 (β_{20}, level-two intercept for *Certainty* coefficient π_2)[e]	*Crt_Hi*: 1.25 *Crt_Med*: 0.625	*Crt_Hi*: 1.259 *Crt_Med*: 0.646	1.000*** (0.007)	0.311***
π_3 (β_{30}, level-two intercept for *Severity* coefficient π_3)[e]	*Sev_Hi*: 2.00 *Sev_Med*: 1.00	*Sev_Hi*: 2.008 *Sev_Med*: 1.003	1.000*** (0.004)	0.500***
Direct and moderating effects of level-two *RiskAverse* on level-one coefficients $\pi_0 - \pi_3$				
β_{01} (π_0, *IntComply*-intercept)	0.25		0.246*** (0.009)	
β_{11} (π_1, *Celerity*)			0.269*** (0.008)	
B_{21} (π_1, *Certainty*)				
B_{31} (π_1, *Severity*)				
Goodness of fit				
Deviance	NA		12,145	
Parameters	NA		10	

[a] N = 10,800 at vignette level; N = 400 at individual level. We show unstandardized coefficient estimates and robust standard errors in parentheses. Level-one variables are uncentered. Level-two variables are grand-mean centered.
[b] Step two CPE models use these step one β values.
[c] True values are the values used to generate Dataset 4, which is based on Dataset 1. True values equal dimension coefficients for *Celerity*, *Certainty*, and *Severity*. They are multiplied by dimension-level manipulation effect (e.g., *Hi* or *Med*).
[d] **B** values calculated using Equation 1. We took s_x and s_y values from the HLM for Windows' MDM template file. $s_{Celerity}$, $s_{Certainty}$, $s_{Severity}$, and $s_{IntComply}$ equal 0.61, 0.51, 0.82, and 1.64, respectively.
[e] Level-one values are fixed.
*** $p < .001$, ** $p < .01$, * $p < .05$.

[2] The level-one π parameters in our multilevel models correspond to the β parameters from the OLS regression models in Tables 4 and 5.

4.4.1 Centering Effects in Step One

In the next demonstration we compare the effects of uncentered and grand-mean centered level-2 variables on uncentered dummy (0, 1) variables in an RCM analyses of the extended Dataset 4 (Table 8, Models 8-3 and 8-4, respectively).

Table 8. Centering Effects on RCM Analysis of Dummy (0, 1) Variables from Factorial Survey: Step One (Dataset 4)[a]

Variables (coefficients)	True value[b]	Model 8-1	Model 8-2[c]	Model 8-3[c] (*RiskAverse* is uncentered)	Model 8-4 (*RiskAverse* is GMC)
Level-one parameters $\pi_0 - \pi_6$ (corresponding level-two slope intercept $\beta_{00} - \beta_{60}$)					
π_0 (β_{00}, level-two intercept for *IntComply* intercept π_0)		3.674*** (.030)	0.702*** (.021)	-0.024 (.030)	0.702*** (.014)
π_1 (β_{10}, level-two intercept for *Cel_Hi* coefficient π_1)	1.50 [1.50 x 1.00]		2.677*** (.029)	1.497*** (.039)	2.677*** (.015)
π_2 (β_{20}, level-two intercept for *Cel_Med* coefficient π_2)	0.75 [1.50 x .50]		1.324*** (.016)	0.753*** (.027)	1.324*** (.010)
π_3 (β_{30}, level-two intercept for *Crt_Hi* coefficient π_3)	1.25 [1.25 x 1.00]		1.259*** (.009)	1.259*** (.009)	1.259*** (.009)
π_4 (β_{40}, level-two intercept for *Crt_Med* coefficient π_4)	0.625 [1.25 x 0.50]		0.646*** (.009)	0.646*** (.009)	0.646*** (.009)
π_5 (β_{50}, level-two intercept for *Sev_Hi* coefficient π_5)	2.00 [2.00 x 1.00]		2.008*** (.009)	2.008*** (.009)	2.008*** (.009)
π_6 (β_{60}, level-two intercept for *Sev_Med* coefficient π_6)	1.00 [2.00 x 0.50]		1.003*** (.009)	1.003*** (.009)	1.003*** (.009)
Direct and moderating effects of *RiskAverse* (Model 8-3) or *RiskAverse*$_{GMC}$ (Model 8-4) on level-one coefficients $\pi_0 - \pi_6$					
β_{01} (π_0, *IntComply*-intercept)	0.25			0.248*** (.009)	0.248*** (.009)
β_{11} (π_1, *Cel_Hi*)	0.40 [0.40 x 1.00]			0.403*** (.012)	0.403*** (.012)
β_{21} (π_2, *Cel_Med*)	0.20 [0.40 x 0.50]			0.195*** (.009)	0.195*** (.009)
β_{31} (π_3, *Crt_Hi*)[d]	0.00				
β_{41} (π_4, *Crt_Med*)[d]	0.00				
β_{51} (π_5, *Sev_Hi*)[d]	0.00				
β_{61} (π_6, *Sev_Med*)[d]	0.00			-0.005** (.001)	-0.004** (.001)
Goodness of fit					
Deviance	NA	40,780.3	15,507.9	11,953.7	11,953.7
Number of parameters	NA	3	9	17	17
Δ Deviance from previous model			-25,272.4***	-3,554.2***	N/A
Δ Deviance from Model 8-2				-3,554.2***	-3,554.2***

[a] N = 10,800 at vignette level; N = 400 at individual level. We show unstandardized coefficient estimates and robust standard errors in parentheses. Level-one variables are uncentered. Model 8-4 is equivalent to Model 6-4.
[b] True values are the values used to generate Dataset 4, which was based on Dataset 1. We show calculations for true values (i.e., dimension coefficients for *Celerity*, *Certainty*, and *Severity* multiplied by dimension-level manipulation effect *Hi* or *Med*) in square brackets.
[c] Values for β_{00} - β_{06} match those of unreported OLS regressions analysis.
[d] Level-one coefficient is fixed.
*** $p < .001$, ** $p < .01$, * $p < .05$.

Note that the "correct" results (i.e., those that match the coefficients used to generate Dataset 4; Appendix D, Table D2) are displayed in Model 8-3, in which *RiskAverse* is uncentered. Cross-level effects in Model 8-3 are based on *RiskAverse*'s intercept, where *RiskAverse* = 0 (i.e., there is no level-2 moderating effect). The moderating variable in Model 8-4, on the other hand, is $RiskAverse_{GMC}$ (i.e., in which the *RiskAverse* value is centered on $\mu_{RiskAverse}$, the grand mean of *RiskAverse*; thus, $RiskAverse_{GMC} = RiskAverse - \mu_{RiskAverse}$). Cross-level effects in Model 8-4 are based on $RiskAverse_{GMC}$'s intercept, where $RiskAverse_{GMC}$ = 0.00 but its effect is equivalent to $\mu_{RiskAverse}$ = 2.925 (Appendix D, Table D4). Thus, *RiskAverse*'s cross-level effects in Model 8-4 are greater than those in Model 8-3; so too are the values and rankings of their subsequent products (i.e., their level-one π and level-two β coefficients). Therefore, *RiskAverse*'s effects are commingled with those of *Cel_Hi* and *Cel_Med* in their respective π_1 and π_2 estimates, which a single-level step one regression cannot separate. This commingling will become important later on when we examine centering effects in step two of the CPE process.

The question that now confronts the researcher is which values for π_1 and π_2 (i.e., the coefficients for *Cel_Hi*, and *Cel_Med*, respectively, in Model 8-3 vs. 8-4) are "correct". The "correctness" of the answer depends on the research question because both results offer useful though different interpretations. If the researcher is interested in ranking *Celerity*, *Certainty*, and *Severity* without *RiskAverse*'s moderating influence on *Celerity*, then Model 8-3 is appropriate because the uncentered *RiskAverse* parcels out its effects in the calculation of π_1 and π_2. On the other hand, if the researcher is interested in ranking *Celerity*, *Certainty*, and *Severity* based on a subject who exhibits an "average" amount of *RiskAverse*—and its corresponding moderating influence on *Celerity*—then Model 8-4 is appropriate because it accommodates an "average" amount of *RiskAverse* via the grand-mean centered $RiskAverse_{GMC}$.

The difference between the two sets of coefficients can be explained mathematically through the formula for calculating π_1, the coefficient for *Cel_Hi*, in Models 8-3 and 8-4 (see Equation 2).

$$\pi_1 = \beta_{10} + \beta_{11} RiskAverse + r_1, \tag{2}$$

where β_{10} represents the "average value" of the π_1 slope across level-two units; β_{11} represents the slope between *RiskAverse* and π_1, and r_1 represents the level-two residuals.

π_1 can be calculated as follows. Model 8-3 shows β_{10} = 1.497, β_{11} = 0.403, and the uncentered value of *RiskAverse* = 0.00. We can further simplify Equation 2 by assuming r_1 = 0 (i.e., the mean level-two residual value). Therefore:

$$\pi_1 = 1.497 + 0.403(0.000) - (0.000) \tag{3}$$

$$\pi_1 = 1.497 \tag{4}$$

Thus, π_1 equals 1.497 when *RiskAverse* = 0.000, which is within the 95 percent CI of the parameter used to generate Dataset 4 (i.e., 1.500). Similar calculations can be made for Model 8-4; however, Model 8-4 is based on $RiskAverse_{GMC}$, which is centered on the "average" level of π_1 (i.e., 2.925). Thus, to compare the π_1 values between Models 8-3 and 8-4, we will need to *subtract* the average level of *RiskAverse* in Equation 5:

$$\pi_1 = 2.677 + 0.403(-2.925) - (0.000) \tag{5}$$

$$\pi_1 = 1.498 \tag{6}$$

Equations 3-4 and 5-6 show Models 8-3 and 8-4 are equivalent (within rounding error). π_1 in Equation 6 equals 1.498 when *RiskAverse* = 0.000 (i.e., $RiskAverse_{GMC}$ = -2.925), which is within the 95 percent CI of the parameter used to generate Dataset 4 (i.e., 1.500). Equations 3-4 and 5-6 can be similarly applied to π_2, the coefficient for *Cel_Med*. The important point here is that the researcher must understand level-two centering effects on the β_{ij} outcomes generated in step one of the CPE process and choose that model that better fits the underlying research question. This decision is important because, as we demonstrate next, it affects the ranking of individual-level factors in step two.

4.4.2 Centering Effects in Step Two

Table 9 displays two types of centering effects in step two.

Table 9. Centering Effects on RCM Analyses of Factorial Surveys: Step Two (Dataset 4)[a]

Model	True value[b]	Model 9-1			Model 9-2		
Level-one centering		Uncentered					
CPE process		Step one inputs[c]	Step two results		Step one inputs[c]	Step two results	
Level-two *RiskAverse* centering		Uncentered	Uncentered		GMC	GMC	
Coefficient		Model 8-3 β	β	**B**[d]	Model 8-4 β	β	**B**[e]
Level-one intercept and independent variables (coefficients) $\pi_0 - \pi_3$							
π_0 (β_{00}, level-two intercept for *IntComply* intercept π_0)	0.00	N/A	-0.032 (0.032)		N/A	0.702*** (0.013)	
π_1 (β_{10}, level-two intercept for *Celerity* coefficient π_1)	*Cel_Hi*: 1.50 *Cel_Med*: 0.75	*Cel_Hi*: 1.497 *Cel_Med*: 0.753	1.000*** (0.026)	0.372***	*Cel_Hi*: 2.677 *Cel_Med*: 1.324	1.000*** (0.005)	0.665***
π_2 (β_{20}, level-two intercept for *Certainty* coefficient π_2)[f]	*Crt_Hi*: 1.25 *Crt_Med*: 0.625	*Crt_Hi*: 1.259 *Crt_Med*: 0.646	1.009*** (0.018)	0.314***	*Crt_Hi*: 1.259 *Crt_Med*: 0.646	1.000*** (0.007)	0.311***
π_3 (β_{30}, level-two intercept for *Severity* coefficient π_3)[f]	*Sev_Hi*: 2.00 *Sev_Med*: 1.00	*Sev_Hi*: 2.008 *Sev_Med*: 1.003	1.003*** (0.011)	0.502***	*Sev_Hi*: 2.008 *Sev_Med*: 1.003	1.000*** (0.004)	0.500***
Direct and moderating effects of level-two *RiskAverse* on level-one coefficients $\pi_0 - \pi_3$							
β_{01} (π_0, *IntComply*-intercept)	0.25		0.248*** (0.010)			0.250*** (0.010)	
β_{11} (π_1, *Celerity*)			0.269*** (0.008)			0.151*** (0.004)	
β_{21} (π_2, *Certainty*)			-0.003 (0.006)			-0.003 (0.006)	
β_{31} (π_3, *Severity*)			-0.001 (0.004)			-0.001 (0.004)	
Goodness of fit							
Deviance	NA		12,143			12,130	
Parameters	NA		19			19	

[a] N = 10,800 at vignette level; N = 400 at individual level. We show unstandardized coefficient estimates and robust standard errors in parentheses.
[b] True values are the values used to generate Dataset 4, which was based on Dataset 1. True values equal dimension coefficients for *Celerity*, *Certainty*, and *Severity*. They are multiplied by dimension-level manipulation effect (e.g., *hi* or *med*).
[c] These step one β values are used as inputs to step two CPE models.
[d] **B** values calculated using Equation 1. We took s_x and s_y values from the HLM for Windows' MDM template file for Model 9-1. $s_{Celerity}$ and $s_{IntComply}$ equal 0.61 and 1.64, respectively.
[e] **B** values calculated using Equation 1. We took s_x and s_y values from the HLM for Windows' MDM template file for Model 9-2. $s_{Celerity}$ and $s_{IntComply}$ equal 1.09 and 1.64, respectively.
[f] Level-one values are fixed.
*** $p < .001$, ** $p < .01$, * $p < .05$.

The first effect involves the way differences in the level-one π values in Model 8-3 versus Model 8-4 affect calculations of the standardized **B** coefficients used to rank variables in CPE step two. Part of the difference concerns which π values one chooses to calculate the standardized **B** coefficients. The second effect involves the way differences in the centering of the level-two variable *RiskAverse* across those models creates differences in the coefficients of the aggregate variables *Celerity*, *Certainty*, and *Severity* (i.e., π_1, π_2, and π_3, respectively). A researcher must consider these different moderating influences when recalibrating the dimension-level intervals in CPE of step two. We demonstrate these CPE calculations in Models 9-1 and 9-2, which are based on the unstandardized π coefficients from Models 8-3 and 8-4, respectively. We calculated the values in Table 9 using HLM for Windows version 7.01 using full maximum likelihood estimation. HLM for Windows does not calculate standardized **B** values, so we calculated them manually using Equation 1 and output in the HLM for Windows MDM files.

The most relevant statistics for the purposes of this demonstration are the values of the **B** coefficient for the step two variable *Celerity*. Based on Equation 1, the value for **B** in Model 9-1 is **B** = $(\beta * s_x)/s_y$ = (1.000 * 0.61)/1.64 = 0.372; for Model 9-2, **B** = (1.000 * 1.09)/1.64 = 0.665. These calculations show that the difference in the **B** values can be traced to their different s_x values (i.e., 0.61 and 1.09 for Models 9-1 and 9-2, respectively). The differences in s_x arise because of the way *RiskAverse* is centered in the two models and the corresponding differences in moderation effects of *RiskAverse* on the *Cel_Med* and *Cel_Hi* relationships with *IntComply* (i.e., 0.00 in Model 9-1; 2.925 in Model 9-2). These centering differences create different step one β values across the two models (i.e., 0.753 for *Cel_Med* and 1.497 for *Cel_Hi* in Model 9-1; 1.324 for *Cel_Med* and 2.677 for *Cel_Hi* in Model 9-2). Since the ranges of the *Cel_Med* and *Cel_Hi* β values are smaller in Model 9-1 than in Model 9-2, the standard deviations of the aggregate *Celerity* variables in the CPE process (i.e., s_x) differs correspondingly (i.e., 0.61 and 1.09 in Models 9-1 and 9-2, respectively). The thrust of these differences in **B** calculations is that *Celerity* is ranked lower than *Severity* in Model 9-1 but higher than *Severity* in Model 9-2. Again, the different results arise due to the difference of the nature of the targeted reference subject. In Model 9-1, the reference subject has a *RiskAverse* value of 0.00, while the reference subject in Model 9-2 is a subject with an "average" level of *RiskAverse* equal to 2.925.

As in Models 8-3 versus 8-4, the researcher must now address the question of which ranking (i.e., Model 9-1 vs. 9-2) is "correct"? Again, it depends on the research question. If the researcher is interested in ranking *Celerity*, *Certainty*, and *Severity* without *RiskAverse*'s moderating influence on *Celerity*, then Model 9-1 is appropriate *if the β inputs from Model 8-3 are used*. On the other hand, if the researcher is interested in ranking *Celerity*, *Certainty*, and *Severity* based on the "average" subject (i.e., with an "average" amount of *RiskAverse*'s moderating influence on *Celerity* factored in), then then Model 9-2 is appropriate *if the β inputs from Model 8-4 are used*. Model 9-2 offers results that are more consistent for those subjects with an "average" level of *RiskAverse*. Again, the important point here is that the researcher must understand the effects of centering on the ranking of individual-level factors in CPE.

A second effect concerns the change in values for the *IntComply*-intercepts (π_0) across Models 9-1 and 9-2. The π_0 for Model 9-1, which uses uncentered *RiskAverse*, equals -0.032, while the π_0 for Model 9-2, which uses $RiskAverse_{GMC}$, equals 0.702. The increase from -0.032 in Model 9-1 to 0.702, which equals 0.734, arises due to the equivalency of the moderating effects of $RiskAverse_{GMC}$ = 0.00 and $\mu_{RiskAverse}$ = 2.925 in Model 9-2 (Equations 7-8):

$$\beta_{01} * RiskAverse_{GMC} = \beta_{01} * \mu_{RiskAverse} \tag{7}$$

$$\beta_{01} * \mu_{RiskAverse} = 0.250 * 2.925 = 0.731 \approx 0.734 \tag{8}$$

A third effect concerns the change in values for the β_{11} coefficients from Model 9-1 to Model 9-2 (i.e., 0.269 and 0.151, respectively). At first glance, it seems incongruous that the β_{11} coefficient would decrease from Models 9-1 to 9-2 given that the value of *RiskAverse* increases from 0.00 in the uncentered Model 9-1 to 2.925 in the grand-mean centered Model 9-2. However, one must remember that the *Cel_Hi* and *Cel_Med* inputs increase across the two models (e.g., *Cel_Hi* increases from 1.497 to 2.677). These latter increases produce an equalizing effect, which we can demonstrate by multiplying the *Cel_Hi* and *Cel_Med* inputs by their corresponding *RiskAverse* values. For example, the product of the *Cel_Hi* input and β_{11} in Model 9-1 equals 1.497 x 0.269 = 0.403, which is roughly equal to that of Model 9-2 (i.e., 2.677

x 0.151 = 0.404). Thus, the ultimate effect on the dependent variable *IntComply* is equivalent across the two models regardless of whether the model is uncentered or grand-mean centered.

It is important that centering choices be applied consistently across steps one and two. We describe some effects of inconsistent centering in step two of the CPE procedure in Appendix F.

4.4.3 Other Centering Points

Centering at the grand mean or at the zero-point uncentered data are not the only two points at which data can be centered. Recentering factorial survey data is useful when 0.00 is outside the data range (e.g., in a 1-5 Likert-type scale or in respondents' age ranges). In such circumstances, the data should be recoded so that the intercept will be placed at a meaningful point in the data range. For example, one could rescale data from a 1-5 Likert-type scale on a 0-4 range (e.g., each "1" is recoded as a "0" and each "5" is recoded as a "4"). One could similarly rescale respondents' ages (e.g., if the minimum age in the sample were 22 years, then one would reduce each age value by 22). Rescaling would make the regression and RCM intercepts meaningful because they would be placed in the data range at the lowest point (i.e., at the 0th percentile).

Researchers can also demonstrate how the relationships between the level-one dependent and independent variables vary across differing values of the level-two variable (e.g., positive, significant moderation effects at high levels of the level-two variable, but non-significant moderating effects at low levels of the level-two variable). Calculating and plotting the slopes of such changes is particularly useful in presentations and publications because they visually demonstrate how slopes change across different values of the level-two variable. Hox (1995, 2002) recommends the 25th, 50th (median), and 75th percentiles as recentering points.

5 Discussion

Our tutorial makes several contributions regarding factor ranking in factorial survey analyses. First, we demonstrate and explicate the usefulness of multilevel modeling. We discuss the value of RCM in estimating variations in rates of change in vignette-level coefficients across individuals. We show that these estimates are more accurate than those of OLS regression, the method of choice for most applications of Rossi and Anderson's (1982) two-step method. We also explain how multilevel modeling accounts for these effects in nested, hierarchical data (e.g., vignettes nested within individuals). We hope our demonstrations promote wider usage of multilevel modeling in factorial survey analysis.

Second, we describe the effects of centering decisions in multilevel analysis. Our demonstration is particularly helpful given the plethora of recommendations for grand-mean centered over uncentered "raw metric" data (e.g., Enders & Tofighi 2007; Hofmann & Gavin 1998; Raudenbush & Bryk, 2002). We agree that grand-mean centering does have several computational and interpretive advantages and that one may prefer it over uncentered data in many circumstances. Grand-mean and group-mean centering level-one variables may be particularly useful for those methods that incorporate continuous level-one independent variables (e.g., policy capturing). Both uncentered and grand-mean centered level-two variables are useful, but their use must correspond to one's interests to avoid compromising the validity and/or interpretability of one's work. Researchers should, therefore, understand how centering affects the estimation, interpretation, and ranking of vignette dimension coefficients.

Third, we demonstrate the adverse effects of unequal dimension intervals and manipulation effect sizes on the ranking of dummy (0, 1) variables in CPE. Such rankings may be unreliable when unequal manipulation effect sizes across the survey's dimensions produce unequal effects on *Y* that can be confounded with the effects of the dimensions in the factorial design. Of course, that does not mean one should necessarily avoid qualitative statements. Indeed, they can be useful when their meanings are clear, unambiguous, realistic, and respondents can easily grasp them. However, achieving these goals can be challenging. Researchers should justify their dimension manipulation choices by relying on current practice, economic costs, or other organizational factors to validate their decisions (Dallal, 2001). They can frame choices in a context of practice (e.g., how a fixed set of resources can best be deployed to reach a targeted goal given a fixed budget or how an organization should invest in various ethical strategies to optimally increase compliance). Input from pilot tests and knowledgeable, impartial third-party experts can help mitigate unrealistic dimension manipulation effect sizes. Such practical foundations can increase the validity of the research design.

Researchers new to multilevel methods typically face challenges in interpreting the meaning of the different statistics. Given that most of our examples and discussion revolve around the effects on various coefficients (e.g., the ranking of standardized **B** coefficients in Table 9), the best place to start may be with the traditional meaning of a regression coefficient. Cohen and Cohen (1983, p. 42) state a regression coefficient "represents the rate of change in *Y* units per *X* unit" (i.e., the dependent and independent variables, respectively;). One can similarly apply this definition in random coefficients modeling (e.g., the level-one parameters are the outcomes of level-two regression equations).

Our results demonstrate how multilevel analyses and techniques allow researchers to model and test nuances in the way an individual's characteristics influence relationships between factors represented in vignettes. In our demonstrations, we examine information security-related examples (e.g., the extent to which an individual's aversion to risk affects the relationships between their attitudes about punishment and their intended compliance behaviors). Future research could investigate other IS topics as well, some of which we suggest in Section 5.2. Researchers could also apply our results to other vignette-oriented techniques such as policy capturing, which uses quantitative rather than qualitative factors. Table 10 summarizes our recommendations.

Table 10. Summary of Recommendations

Recommendation	Justification
Use current practice, economic costs, and/or other relevant factors when designing dimension manipulation effect sizes. Pilot test values for manipulation effect sizes. Use quantified variables when appropriate.	The participants choose the values of independent variables in most survey analyses, but the values of those variables in factorial survey analyses are forced on them. As a result, scale ranges in traditional survey analyses depend on the choices of the participant pool, while those in factorial survey analysis depend on the manipulation choices of the research designer(s). Therefore, researchers should apply current practices, economic costs, or other organizational or environment factors to design valid, realistic alternatives for their sample. Manipulations that are too wide may include options that are rare or infeasible, while those that are too narrow may limit the independent variable's effect size. Qualitative statements should be pilot tested for manipulation effect size and not just clarity and readability.
Frame dimension rankings in terms of applicable research or practitioner goals.	Dimension importance is determined by the underlying organizational, environmental, or economic factors, which CPE does not account for. When discussing dimension importance, consider recommendations such as how and/or where resources can be best deployed to reach a targeted goal (e.g., given a fixed budget, to what extent should we invest in various sanction severity, certainty, and celerity strategies to most economically increase compliance?).
Center level-two variables consistently across steps one and two of the CPE process.	Mixing centering methods for a level-two variable across CPE steps (e.g., grand-mean centering a level-two variable in step two that has been left uncentered in step one) produces statistical results that can lead to erroneous conclusions (Appendix F).
Grand-mean center individual-level (i.e., level-two) variables if one seeks to estimate level-one coefficients around the average level of the level-two variable.	Grand-mean centering level-two variables produces level-one intercepts that are centered on the average characteristics of individuals in the sample pool, which often provides statistics that are more easily interpreted because they refer to an individual with an average levels of given characteristics in the sample pool rather than an atypical individual who scores zero on all measures. Grand-mean centering is particularly useful when the zero value of a level-two variable is not meaningful or outside the range of values in the sample pool (e.g., an age of 0).

5.1 Limitations

Like other research, our study has limitations. First, we use normally distributed datasets. Future research should explore the effects of non-normally distributed data. Censored data may be a particularly fruitful subject of research. Second, we use simulated data. In particular, our analyses assume that participants can evaluate all 27 vignettes without fatigue or cognitive biases. We caution readers that such assumptions may be unrealistic with real test subjects. Cognitive biases may be particularly interesting because of their strong association with decision making and information systems use.

5.2 Future Research

Several opportunities for future research remain. First, future IS research that uses multilevel factorial surveys could explore the value of rank-ordered beta weights. Such rankings have been used in a variety

of non-IS research domains and can be extended to IS. For example, one could extend Shlay et al.'s (2005) use of beta weights for ranking predictors of summative judgments of childcare quality to explore judgments of IS quality. Harris and Weistroffer (2009) describe several relevant predictors of system success judgments, some of which one could conceptualize as individual-level factors (e.g., user expertise and task competence) and others as vignette-level factors (e.g., user participation and involvement). One could make a second extension from Hunter and McClelland's (1991) ranking of predictors of perceived seriousness of student-student sexual harassment to similar judgments about the perceived seriousness of IS-related behaviors (e.g., cyberloafing, Lim & Teo, 2005; IS consultant wrongdoing, Ayers & Kaplan, 2005). One could also apply perceived seriousness to IS education, particularly in regards to perceived seriousness of academic dishonesty offenses. Using Broeckelman-Post's (2008) examination of academic dishonesty as a guide, researchers could design multilevel studies around individual-level predictors such as a student's understanding of academic expectations, domestic versus international status, and whether students see cheating as a way to get ahead or as a way "for staying on the right side of the growing wage gap between the 'haves' and the 'have-nots'" (Brockelman-Post, 2008, p. 206, referencing Callahan, 2004). Vignette-level predictors could include the number of times an instructor has discussed academic dishonesty policies and the amount of feedback an instructor has given about source citation. One could make a third extension from Alves' (1982) rankings of predictors of distributive justice judgments. Specifically, one could extend this line of research to investigations of distributive justice judgments in knowledge sharing (e.g., Fang & Chiu, 2010), IS security policy compliance (e.g., Xue, Liang, & Wu, 2011), or other IS-related concerns.

Multilevel modeling offers a sound methodological tool for analyzing theoretically important relationships that involve individual-level factors in factorial surveys. However, past factorial survey research in the IS domain has limited its use primarily to controlling for the effects of demographic variables such as age and gender (e.g., Guo et al, 2011; Vance et al., 2015). Expanding the scope of multilevel modeling beyond this limited role can facilitate IS researchers' efforts to explore relationships involving other, more theoretically relevant individual-level factors such as perceptions of supervisors' and coworkers attitudes and behaviors (Webster & Trevino, 1995), a subject's overall job characteristics (Van de Ven & Ferry, 1980, cited in Webster & Trevino, 1995), or perceptions about an organization's social media use policy (Trinkle et al, 2014). Moreover, one can model such attitudes and perceptions more effectively as continuous direct or moderating cross-level variables than as median-split groups in structural equation modeling (e.g., Moores & Chang, 2006). The entrepreneurship literature may be particularly informative in this regard. For example, several individual-level factors studied in the entrepreneurship literature (e.g., trust, Patzelt & Shepard, 2008; self-efficacy, Mitchell & Shepherd, 2010) have also been studied in IS research (e.g., Gefen, Karahanna, & Straub, 2003; Agarwal, Sambamurthy, & Stair, 2000). Future IS research should take advantage of these opportunities.

Future research could also address how to incorporate mediating psychological states (e.g., O'Keefe, 2003) in factorial survey analyses (e.g., in Brunswik's (1952) lens model of behavior; Hammond, 1966; Dalal, Diab, Balzer, & Doherty, 2010). Such research could extend factorial surveys by including psychological states in this methodology.

6 Conclusion

Our analyses of manipulation, multilevel, and centering effects indicate that researchers should thoughtfully plan for and carefully execute coding proportional to effect procedures to rank dimensions by importance in factorial survey analysis. Researchers should understand how multilevel modeling and centering affect and improve CPE factor rankings and ensure that their methodological decisions align with their research goals.

Qualitative statements may elicit manipulation effects that are too pronounced or that may include rare or infeasible options. On the other hand, qualitative statements that are too limited may fail to elicit realistic responses in the dependent variable. As such, researchers should base their qualitative vignette dimensions with relevant, realistic organizational examples. Our results indicate that quantified statements should replace qualitative statements and their dummy (0, 1) variables when possible.

We hope our demonstrations and recommendations increase the validity of factorial survey research and contribute to best practices for future research.

Acknowledgments

We thank the editors, reviewers, and copyeditor for their insights, assistance, and guidance on earlier versions of this manuscript. We also thank the participants in The 2012 Dewald Roode Workshop on Information Systems Security Research, IFIP WG8.11/WG11.13, for their helpful feedback on an earlier presentation of this research.

References

Agarwal, R., Sambamurthy, V., & Stair, R. M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy—an empirical assessment. *Information Systems Research*, *11*(4), 418-430.

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351-371.

Aiman-Smith, L., Scullen, S. E., & Barr, S. H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, *5*(4), 388-414.

Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly*, *42*(1), 93-104.

Alves, W. M. (1982). Modeling distributive justice judgments. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 205-234). Beverly Hills, CA: Sage.

Ayers, S., & Kaplan, S. E. (2005). Wrongdoing by consultants: An examination of employees' reporting intentions. *Journal of Business Ethics*, *57*(2), 121-137.

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. R. (2013). Don't make excuses! Discouraging neutralization to reduce IT policy violation. *Computers & Security*, *39*, 145-159.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 1017-1041.

Bouwman, H., & van de Wijngaert, L. (2009). Coppers context, and conjoints: A reassessment of TAM. *Journal of Information Technology*, *24*(2), 186-201.

Boyle, R. P. (1970). Path analysis and ordinal data. *American Journal of Sociology*, *75*(4), 461-480.

Broeckelman-Post, M. A. (2008). Faculty and student classroom influences on academic dishonesty. *IEEE Transactions on Education*, *51*(2), 206-211.

Brunswik, E. (1952). The conceptual framework of psychology. *International Encyclopedia of Unified Science*, *1*(10). Chicago, IL: The University of Chicago Press.

Burton-Jones, A., & Gallivan, M. J. (2007). Toward a deeper understanding of system usage in organizations: A multilevel perspective. *MIS Quarterly*, *31*(4), 657-679.

Callahan, D. (2004). *The cheating culture: Why more Americans are doing wrong to get ahead*. Orlando, FL: Harcourt.

Chen, G. (2014). When and how to center a variable? *National Institute for Mental Health*. Retrieved from https://sscc.nimh.nih.gov/sscc/gangc/centering.html

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & Security*, *32*, 90-101.

Dalal, D. K., Diab, D. L., Balzer, W. K., & Doherty, M. E. (2010). The lens model: An application of JDM methodologies to IOOB practice. *Industrial and Organizational Psychology*, *3*(4), 424-428.

Dallal, G. E. (2001). *Which predictors are more important?* Retrieved from http://www.jerrydallal.com/LHSP/importnt.htm

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, *20*(1), 79-98.

Easton, V. J., & McColl, J. H. (1997). *Statistics glossary v1.1*. Retrieved from http://www.stats.gla.ac.uk/steps/glossary/

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue, *Psychological Methods*, *12*(2), 121-138.

Fang, Y. H., & Chiu, C.-M. (2010). In justice we trust: Exploring knowledge-sharing continuance intentions in virtual communities of practice. *Computers in Human Behavior*, *26*(2), 235-246.

Gefen, D., Karahanna, E., & Straub, D.W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51-90.

Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E. (2011). Understanding non-malicious security violations in the workplace: A composite behavior model. *Journal of Management Information Systems*, *28*(2), 203-236.

Hammond, K. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. E. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 15-80). New York: Holt, Rinehart, and Winston.

Harrington, S. J. (1996). The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. *MIS Quarterly*, *20*(3), 257-278.

Harris, M. A., & Weistroffer, H. R. (2009). A new look at the relationship between user involvement in systems development and systems success. *Communications of the Association for Information Systems*, *24*, 739-756.

Herzog, S. (2003). The relationship between public perceptions of crime seriousness and support for plea-bargaining practices in Israel: A factorial survey approach. *Journal of Criminal Law & Criminology*, *94*(1), 103-131.

Hofmann, D. A., & Gavin M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations, *Journal of Management*, *24*(5), 623-641.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Hox, J., Kreft, I. G. G., & Hermkens, P. L. J. (1991). The analysis of factorial surveys. *Sociological Methods & Research*, *19*(4), 493-510.

Hunter, C., & McClelland, K. (1991). Honoring accounts for sexual harassment: A factorial survey analysis. *Sex Roles*, *24*(11/12), 725-752.

Johnston, A. C., Warkentin, M., McBride, M., & Carter, L. (2016). Dispositional and situational factors: Influences on IS security policy violations. *European Journal of Information Systems*, *25*(3), 231-251.

Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, *5*(4), 337-361.

Kelloway, E. K. (n.d.). Structural equation modeling with MPlus. *CN Centre for Occupational Health & Safety*. Available at ohpsychology.ca/wp-content/uploads/2011/02/Day-1.pptx

Klein, K. J., & Kozlowski, S. W. J. (Eds.). (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.

Kreft, I. G. G. (1994). Multilevel models for hierarchically nested data: Potential applications in substance abuse prevention research. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (pp. 142-183). Rockville, MD: National Institute on Drug Abuse.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30(1), 1-21.

Kriauciunas, A., Parmigiani, A., & Rivera-Santos, M. (2011). Leaving our comfort zone: Integrating established practices with unique adaptations to conduct survey-based strategy research in nontraditional contexts. *Strategic Management Journal*, *32*(9), 994-1010.

Krieger, A. M., Green, P. E., & Umesh, U. N. (1998). Effect of level of disaggregation on conjoint cross validations: Some comparative findings. *Decision Sciences*, *29*(4), 1047-1058.

Lauder, W. (2002). Factorial survey methods: A valuable but under-utilized research method in nursing research? *NT Research*, *7*(1), 35-43.

Lee, J., Jr., Crossler, R. E., & Warkentin, M. (2013). Implications of monitoring mechanisms on bring your own device (BYOD) adoption. In *Proceedings of the 2013 International Conference on Information Systems.*

Lim, V. K., & Teo, T. S. (2005). Prevalence, perceived seriousness, justification and regulation of cyberloafing in Singapore: An exploratory study. *Information & Management*, *42*(8), 1081-1093.

Maas, C. J. M., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*(3), 427-440.

Mitchell, J. R., & Shepherd, D. A. (2010). To thine own self be true: Images of self, images of opportunity, and entrepreneurial action. *Journal of Business Venturing*, *25*(1), 138-154.

Mithas, S., Ramasubbu, N., Krishnan, M. S., & Fornell, C. (2006-7). Designing Web sites for customer loyalty across business domains: A multilevel analysis. *Journal of Management Information Systems*, *23*(3), 97-127.

Moores, T., & Chang, J. (2006). Ethical decision making in software piracy: Initial development and test of a four-component model. *MIS Quarterly*, *30*(1), 167-180.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.

O'Fallon, M., & Butterfield, K. (2005). A review of the empirical ethical decision-making literature: 1996-2003. *Journal of Business Ethics*, *59*(4), 375-413.

O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory*, *13*(3), 251-274.

Ongena, G., van de Wijngaert, L., & Huizer, E. (2013). Exploring determinants of early user acceptance for an audio-visual heritage archive service using the vignette method. *Behaviour & Information Technology*, *32*(12), 1216-1224.

Patzelt, H., & Shepherd, D. A. (2008). The decision to persist with underperforming alliances: The role of trust and control." *Journal of Management Studies*, *45*(7), 1217-1243.

Podsakoff, P. M., MacKenzie, S. B., & Lee, J.-Y. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.).Thousand Oaks, CA: Sage.

Rossi, P. H., & Anderson, A. B. (1982). The factorial approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgments: The factorial survey approach* (pp. 15-67). Beverly Hills, CA: Sage.

Shlay, A. B., Tran, H., Weinraub, M., & Harmon, M. (2005). Teasing apart the child care conundrum: A factorial survey analysis of perceptions of child care quality, fair market price and willingness to pay by low-income, African American parents. *Early Childhood Research Quarterly*, *20*(4), 393-416.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Steckel, J., DeSarbo, W. S., & Mahajan, V. (1991). On the creation of acceptable conjoint analysis experimental designs. *Decision Sciences*, *22*(2), 435-442.

Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 41-74). Amsterdam: Elsevier Science Publishers.

Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management*, *30*(2), 165-187.

Trevino, L. K., & Victor, B. (1992). Peer reporting of unethical behavior: A social context perspective. *Academy of Management Journal*, *35*(1), 38-64.

Trinkle, B. S., Crossler, R. E., & Warkentin, M. (2014). I'm game, are you? Reducing real-world security threats by managing employee activity in virtual environments. *Journal of Information Systems*, *28*(2), 307-327.

Van de Ven, A. H. & Ferry, D. L. (1980). *Measuring and assessing organizations*. New York, NY: Wiley.

Vance, A., Lowry, P. B., & Eggett, D. (2013). Using accountability to reduce access policy violations in information systems. *Journal of Management Information Systems*, *29*(4), 263-289.

Vance, A., Lowry, P. B., & Eggett, D. (2015). A new approach to the problem of access policy violations: Increasing perceptions of accountability through the user interface. *MIS Quarterly*, *39*(2), 345-366.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, *38*(2), 505-520.

Webster, J., & Trevino, L. K. (1995). Rational and social theories as complementary explanations of communication media choices: Two policy-capturing studies. *Academy of Management Journal*, *38*(6), 1544-1572.

Willison, R., Warkentin, M., & Johnston, A. C. (2016). Examining employee computer abuse intentions: Insights from justice, deterrence, and neutralization perspectives. *Information Systems Journal*.

Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, *80*(391), 513-524.

Xue, Y., Liang, H., & Wu, L. (2011). Punishment, justice, and compliance in mandatory IT settings. *Information Systems Research*, *22*(2), 400-414.

Appendix A: An Example of a Factorial Survey Vignette

Factorial surveys typically embed the antecedent variables as statements in vignettes (or scenarios) that are displayed to the respondent. Vignettes often state the behavior of a character under orthogonally manipulated conditions; the respondents are asked the degree to which they agree that they would act as the vignette character acted in the same situation. One recent example (from Johnston et al., 2016, p. 249) appears below (in italics). We take the bold sentences from the set of manipulations below. Note the action in the penultimate sentence—the respondent will agree or disagree that they would do the same (as the dependent variable).

Sample vignette (from a set of vignette versions)

Joe has just collected sensitive customer data for his company, and he wants to take that data home to continue his work. He knows his company requires that he request a password to be issued and applied to all data before taking it out of the office on a USB drive so that it cannot be accessed by an unauthorized individual. Joe has completed the password request procedure before, ***so he is confident he can do it again easily****. Joe believes that without the password, it is* ***not likely that unauthorized people will see the data, but if they do, nothing bad will happen****. Joe believes that the password procedure is* ***effective and prevents unauthorized people from seeing the data****. Regardless, the password procedure takes several minutes, and he needs to leave now, so he skips the procedure.* ***Joe believes his chances of being caught are low, but if caught, the punishment would be minimal****.*

Constructs manipulated in the vignettes

Below are the statements associated with the various levels of each of the situational factors manipulated in the vignettes. The levels are shown in parentheses.

Self-efficacy levels

- *Joe has completed the password request procedure before, but he is not confident he can do it again easily (low)*
- *Joe has completed the password request procedure before, so he is confident he can do it again easily (high)*

Threat vulnerability and Severity

- *Joe believes that, without the password, it is not likely that unauthorized people will see the data, but if they do, nothing bad will happen (low/low)*
- *Joe believes that, without the password, it is not likely that unauthorized people will see the data, but if they do, they may alter or misuse it (low/high)*
- *Joe believes that, without the password, it is likely that unauthorized people will see the data, but if they do, nothing bad will happen (high/low)*
- *Joe believes that, without the password, it is likely that unauthorized people will see the data and if they do, they may alter or misuse it (high/high)*

Sanction Certainty and Severity

- *Joe believes his chances of being caught are low, but if caught, the punishment would be minimal (low/low)*
- *Joe believes his chances of being caught are low, but if caught, the punishment would be severe (low/high)*
- *Joe believes his chances of being caught are high, and if caught, the punishment would be minimal (high/low)*
- *Joe believes his chances of being caught are high, and if caught, the punishment would be severe (high/high)*

Response efficacy

- *Joe believes that the password procedure is not effective and does not prevent unauthorized people from seeing the data (low)*
- *Joe believes that the password procedure is effective and prevents unauthorized people from seeing the data (high).*

Appendix B: Generation of Datasets 1-3

The effect sizes of the dimension levels on the dependent variable *IntComply* are based on two factors. The first is a dummy (0, 1) variable that represents the presence of the dimension level's corresponding statement in the vignette. The second represents the magnitude of the dimension level effect. Our design follows a traditional format "where level implies amount or magnitude" (Easton & McColl, 1997). We use three dimension levels for *Celerity*, *Certainty*, and *Severity* to ensure each dimension is over-identified and linear relationships are assessable. We set the *IntComply*-intercept β_0 to zero. Thus, the equations for dummy (0, 1) variables for the *Celerity* dimension are as follows:

$$Cel_Low_{1iv} = Cel_LowPres_{1iv} * (Cel_LowMag_{1iv} + e_{iv}) \quad \text{(B1)}$$

$$Cel_Med_{1iv} = Cel_MedPres_{1iv} * (cel_MedMag_{1iv} + e_{iv}) \quad \text{(B2)}$$

$$Cel_Hi_{1iv} = Cel_HiPres_{1iv} * (cel_HiMag_{1iv} + e_{iv}), \quad \text{(B3)}$$

where *Cel_LowPres*, *Cel_MedPres*, and *Cel_HiPres* represent the presence of a low-, medium-, or high-level statement, respectively, in a vignette; *Cel_LowMag*, *cel_MedMag*, and *Cel_HiMag* represent the magnitude of the statement effect size regarding low, medium, and high dimension levels, respectively; and e_{iv} represents residuals.

We generated Dataset 1 so that the manipulation effect sizes of the first independent variable *Celerity* ranged from a low value of 0.00 to a high value of 1.00 with a mid-range value of 0.50. We set the *Celerity* coefficient*'s* association with *IntComply* to 1.50. We generated Dataset 2 so that the manipulation effect sizes of the *Celerity* ranged from a low level of 0.00 to a high level of 3.00 with a mid-range level of 1.50. To make the interval scale in Dataset 2 equivalent to that in Dataset 1, we set the *Celerity* coefficient to 0.50 for all three levels (e.g., the *Cel_Hi* * β_1 coefficient in Dataset 1 equals 1.00 * 1.50 = 1.50; for Dataset 2, 3.00 * 0.50 = 1.50). We generated Dataset 3 so that the manipulation effect sizes of *Celerity* ranged across a greater span than that of Dataset 1 (i.e., at a low level of 0.00, a mid-range level of 1.50, and a high level of 3.00). We kept the *Celerity* coefficient at 1.50 for both datasets to demonstrate the effects of the increase in dimension interval size.

Figure B1 displays the R code we used to generate step one data for Datasets 1-3 for Table 2.

```
# R code to generate step one data for Datasets 1-3 in Table 2.
# Datasets will be stored in the working directory in a new folder, "FSM_Datasets".
#
# NOTE: Dummy variables in the manuscript (e.g., Cel_Low, Cel_Med, and Cel_Hi)
# correspond to variables with the prefix "dum" in this code (e.g., dumCelLow,
# dumCelMed, and dumCelHi, respectively).
#
# Which dataset do you want to generate? Modify the next line accordingly.
datasetNumber <- 1     # Change value to 1, 2, or 3.
#
# Celerity constants and beta coefficients vary across Datasets 1-3. Values used depend on Line 4.
if(datasetNumber == 1) {
   celCoeff <- 1.50      # Celerity beta coefficient, Dataset 1.
   celHi <- 1.0          # Relative value for high Celerity level, Dataset 1.
   celMed <- 0.5         # Relative value for medium Celerity level, Dataset 1.
   celLow <- 0.0         # Relative value for low Celerity level, Dataset 1.
 } else if(datasetNumber == 2) {
   celCoeff <- 0.50      # Celerity beta coefficient, Dataset 2.
   celHi <- 3.0          # Relative value for high Celerity level, Dataset 2.
   celMed <- 1.5         # Relative value for medium Celerity level, Dataset 2.
   celLow <- 0.0         # Relative value for low Celerity level, Dataset 2.
 } else if(datasetNumber == 3) {
   celCoeff <- 1.50      # Celerity beta coefficient, Datase 3
   celHi <- 3.0          # Relative value for high Celerity level, Dataset 3.
   celMed <- 1.5         # Relative value for medium Celerity level, Dataset 3.
   celLow <- 0.0         # Relative value for low Celerity level, Dataset 3.
```

```
  } else{
    cat("DatasetNumber in Line 4 is incorrect. Please enter 1, 2, or 3.")
    quit()
  }
# Certainty and Severity beta coefficients and constants do not change
# across Datasets 1-3.
crtCoeff <- 1.25     # Certainty coefficient, Datasets 1-3.
crtHi <- 1.0         # High value for Certainty variable, Datasets 1-3.
crtMed <- 0.5        # Medium value for Certainty variable, Datasets 1-3.
crtLow <- 0.0        # Low value for Certainty variable, Datasets 1-3.
sevCoeff <- 2.00     # Severity coefficient, Datasets 1-3.
sevHi <- 1.0         # High value for Severity variable, Datasets 1-3.
sevMed <- 0.5        # Medium value for Severity variable, Datasets 1-3.
sevLow <- 0.0        # Low value for Severity variable, Datasets 1-3.
#
# Parameters used to generate error values, Datasets 1-3. These do not change
# across Datasets 1-3.
r_sd <- 0.4          # Standard deviation for level-1 residual.

# Generate vignette ID numbers (vignID).
vignID <- seq(from = 1001, to = 11800, by = 1)
#
# Generate subject count numbers (vnCount).
for(i in 1:400){
  if(i <= 1) {
    vnCount <- seq(from = 1, to = 27, by = 1)
  } else {
    vnCount <- c(vnCount, seq(from = 1, to = 27, by = 1))
  }
}
#
# Generate subject (individuals) ID numbers (indivID).
indivID <- c(101, 101)
for(i in 3:27) {
  indivID <- c(indivID, 101)
}
for(i in 102:500) {
  for(j in 1:27) {
    indivID <- c(indivID, i)
  }
}
#
# Calculate residuals.
r <- rnorm(10801, mean=0, sd = r_sd)
#
# The following for loop generates the dummy variables that represent which
# Celerity, Certainty, and Severity statements (i.e., High, Medium, or Low)
# are incorporated in each vignette.
# Initialize arrays in loop; these initial rows will be deleted at end of
# the for loops (Lines 150-171).
rec_count   <- 0
rec_countVector <-0
dumCelHi  <- 0
dumCelMed <- 0
dumCelLow <- 0
dumCrtHi  <- 0
dumCrtMed <- 0
dumCrtLow <- 0
```

```
dumSevHi  <- 0
dumSevMed <- 0
dumSevLow <- 0
for(m in 1:400) {
  for(i in 1:3) {
    for(j in 1:3) {
      for(k in 1:3) {
        rec_count <- rec_count + 1
        rec_countVector <- c(rec_countVector, rec_count)
        if(i == 1) {
          dumCelHi  <- c(dumCelHi, 1)
          dumCelMed <- c(dumCelMed, 0)
          dumCelLow <- c(dumCelLow, 0)
        } else if(i == 2) {
          dumCelHi  <- c(dumCelHi, 0)
          dumCelMed <- c(dumCelMed, 1)
          dumCelLow <- c(dumCelLow, 0)
        } else {
          dumCelHi  <- c(dumCelHi, 0)
          dumCelMed <- c(dumCelMed, 0)
          dumCelLow <- c(dumCelLow, 1)
        }
        if(j == 1) {
          dumCrtHi  <- c(dumCrtHi, 1)
          dumCrtMed <- c(dumCrtMed, 0)
          dumCrtLow <- c(dumCrtLow, 0)
        } else if(j == 2) {
          dumCrtHi  <- c(dumCrtHi, 0)
          dumCrtMed <- c(dumCrtMed, 1)
          dumCrtLow <- c(dumCrtLow, 0)
        } else {
          dumCrtHi  <- c(dumCrtHi, 0)
          dumCrtMed <- c(dumCrtMed, 0)
          dumCrtLow <- c(dumCrtLow, 1)
        }
        if(k == 1) {
          dumSevHi  <- c(dumSevHi, 1)
          dumSevMed <- c(dumSevMed, 0)
          dumSevLow <- c(dumSevLow, 0)
        } else if(k == 2) {
          dumSevHi  <- c(dumSevHi, 0)
          dumSevMed <- c(dumSevMed, 1)
          dumSevLow <- c(dumSevLow, 0)
        } else {
          dumSevHi  <- c(dumSevHi, 0)
          dumSevMed <- c(dumSevMed, 0)
          dumSevLow <- c(dumSevLow, 1)
        }
        # Calculate Marginal Effect (AME) of each dimension level.
        celHiME <- (dumCelHi * celHi * celCoeff)
        celMedME <- (dumCelMed * celMed * celCoeff)
        celLowME <- (dumCelLow * celLow * celCoeff)
        crtHiME <- (dumCrtHi * crtHi * crtCoeff)
        crtMedME <- (dumCrtMed * crtMed * crtCoeff)
        crtLowME <- (dumCrtLow * crtLow * crtCoeff)
        sevHiME <- (dumSevHi * sevHi * sevCoeff)
        sevMedME <- (dumSevMed * sevMed * sevCoeff)
        sevLowME <- (dumSevLow * sevLow * sevCoeff)
```

```
      # Calculate sum of marginal effect (sumMargEffect, the sum of corresponding AMEs)
       sumMargEffect <- celHiME + celMedME + celLowME +
                crtHiME + crtMedME + crtLowME +
                sevHiME + sevMedME + sevLowME
     }
    }
  }
}
# Add total effect size (sumMargEffect) and residual (r) to compute dependent variable, intComply.
intComply <- sumMargEffect + r
#
# Remove first row of "0" values from above vectors (many initialized in Lines 72-82).
rec_countVector <- rec_countVector[-1]
dumCelHi <- dumCelHi[-1]
dumCelMed <- dumCelMed[-1]
dumCelLow <- dumCelLow[-1]
dumCrtHi <- dumCrtHi[-1]
dumCrtMed <- dumCrtMed[-1]
dumCrtLow <- dumCrtLow[-1]
dumSevHi <- dumSevHi[-1]
dumSevMed <- dumSevMed[-1]
dumSevLow <- dumSevLow[-1]
celHiME <- celHiME[-1]
celMedME <- celMedME[-1]
celLowME <- celLowME[-1]
crtHiME <- crtHiME[-1]
crtMedME <- crtMedME[-1]
crtLowME <- crtLowME[-1]
sevHiME <- sevHiME[-1]
sevMedME <- sevMedME[-1]
sevLowME <- sevLowME[-1]
sumMargEffect <- sumMargEffect[-1]
r <- r[-1]
intComply <- intComply[-1]
#
# Create csv file of variable values.
fileContents <- cbind(vignID, vnCount, indivID)
#
# Dependent variable
fileContents <- cbind(fileContents, intComply)
#
# Dummy variables
fileContents <- cbind(fileContents, dumCelHi)
fileContents <- cbind(fileContents, dumCelMed)
fileContents <- cbind(fileContents, dumCelLow)
fileContents <- cbind(fileContents, dumCrtHi)
fileContents <- cbind(fileContents, dumCrtMed)
fileContents <- cbind(fileContents, dumCrtLow)
fileContents <- cbind(fileContents, dumSevHi)
fileContents <- cbind(fileContents, dumSevMed)
fileContents <- cbind(fileContents, dumSevLow)
#
# Manipulation and marginal effects
fileContents <- cbind(fileContents, celHi)
fileContents <- cbind(fileContents, celMed)
fileContents <- cbind(fileContents, celLow)
fileContents <- cbind(fileContents, celCoeff)
fileContents <- cbind(fileContents, celHiME)
```

```
fileContents <- cbind(fileContents, celMedME)
fileContents <- cbind(fileContents, celLowME)
#
fileContents <- cbind(fileContents, crtHi)
fileContents <- cbind(fileContents, crtMed)
fileContents <- cbind(fileContents, crtLow)
fileContents <- cbind(fileContents, crtCoeff)
fileContents <- cbind(fileContents, crtHiME)
fileContents <- cbind(fileContents, crtMedME)
fileContents <- cbind(fileContents, crtLowME)
#
fileContents <- cbind(fileContents, sevHi)
fileContents <- cbind(fileContents, sevMed)
fileContents <- cbind(fileContents, sevLow)
fileContents <- cbind(fileContents, sevCoeff)
fileContents <- cbind(fileContents, sevHiME)
fileContents <- cbind(fileContents, sevMedME)
fileContents <- cbind(fileContents, sevLowME)
#
# Sum of marginal effects (used to check calculations)
fileContents <- cbind(fileContents, sumMargEffect)
#
# Residual
fileContents <- cbind(fileContents, r)
#
# Create step one dataset output file.
# Add “FSM_Datasets” folder to working directory if necessary.
if(file.exists(file.path(getwd(),"FSM_Datasets"))) {
   # If the FSM_Datasets folder exists, do nothing.
} else {
   dir.create(file.path(getwd(),"FSM_Datasets"))
}
if(datasetNumber == 1) {
   write.csv(fileContents, file="FSM_Datasets/R_FigureB1_Dataset1_Step1.csv",
            row.names = FALSE)
} else if(datasetNumber == 2) {
   write.csv(fileContents, file="FSM_Datasets/R_FigureB1_Dataset2_Step1.csv",
            row.names = FALSE)
} else if(datasetNumber == 3) {
   write.csv(fileContents, file="FSM_Datasets/R_FigureB1_Dataset3_Step1.csv",
            row.names = FALSE)
} else {
  cat("DatasetNumber in Line 4 is incorrect. Please enter 1, 2, or 3.")
  quit()
}
```

Figure B1. R Code to Generate Datasets 1-3, Step One, Table 2

Figure B2 displays the SPSS code we used to perform the OLS regression analysis for step one in Table 2.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT intComply
  /METHOD=ENTER dumCelHi dumCelMed dumCrtHi dumCrtMed dumSevHi dumSevMed.
```

Figure B2. SPSS Code for First OLS Regression, Datasets 1-3

Figure B3 displays the R code we used to generate step two (CPE) data for Datasets 1-3 in Table 3.

```
# R code for Table 3, Models 3-1, 3-2, and 3-3.
# Datasets will be stored in the working directory in a new folder, “FSM_Datasets”.
#
# Which model do you wish to use apply the step two (CPE) method?
# Be sure to enter the step one beta coefficients in lines 24 to 29 below.
# Enter:
#   "1" for Model 3-1
#   "2" for Model 3-2
#   "3" for Model 3-3
#
modelNum <- 1 # Model 3-1 is the default choice. Change modelNum as needed.
#
if (modelNum == 1) {
      data <- read.csv(file = "FSM_Datasets/R_FigureB1_Dataset1_Step1.csv")
} else if(modelNum == 2) {
      data <- read.csv(file = "FSM_Datasets/R_FigureB1_Dataset2_Step1.csv")
} else {
      data <- read.csv(file = "FSM_Datasets/R_FigureB1_Dataset3_Step1.csv")
}
#
# Enter unstandardized beta coefficients for the dummy variables from the SPSS
# OLS regression in Table 2.
#
Celerity_High <- 4.509
Celerity_Med <- 2.265
Certainty_High <- 1.254
Certainty_Med <- 0.631
Severity_High <- 1.991
Severity_Med <- 0.987
#
# Generate aggregate variables.
Celerity <- (data[5] * Celerity_High) + (data[6] * Celerity_Med)
Certainty <- (data[8] * Certainty_High) + (data[9] * Certainty_Med)
Severity <- (data[11] * Severity_High) + (data[12] * Severity_Med)
#
# Create csv file that mimics Table10Data.
fileContents <- cbind(data[,1:13], Celerity)
fileContents <- cbind(fileContents, Certainty)
fileContents <- cbind(fileContents, Severity)
colnames(fileContents) <- c("vignID", "count", "indivID", "IntComply",
                            "Cel_Hi", "Cel_Med", "Cel_Low",
                            "Crt_Hi", "Crt_Med", "Crt_Low",
                            "Sev_Hi", "Sev_Med", "Sev_Low",
                            "Celerity", "Certainty", "Severity")
# Create csv output file.
# Add "FSM_Datasets" folder to working directory if necessary.
```

```
if(file.exists(file.path(getwd(),"FSM_Datasets"))) {
   # If the FSM_Datasets folder exists, do nothing.
} else {
   dir.create(file.path(getwd(),"FSM_Datasets"))
}
if (modelNum == 1) {
  write.csv(fileContents, file="FSM_Datasets/R_Model_3-1_Step2CPE.csv",
            row.names = FALSE)
} else if(modelNum == 2) {
  write.csv(fileContents, file="FSM_Datasets/R_Model_3-2_Step2CPE.csv",
            row.names = FALSE)
} else {
  write.csv(fileContents, file="FSM_Datasets/R_Model_3-3_Step2CPE.csv",
            row.names = FALSE)
}
```

Figure B3. R Code for CPE Step Two for Datasets 1-4

Table B1 displays the descriptive statistics of dimension levels from Datasets 1-3 (e.g., *Cel_Hi*, *Cel_Med*, and *Cel_Low*). Table B2 displays the descriptive statistics of the dimensions from Datasets 1-3 (i.e., *Celerity*, *Certainty*, and *Severity*).

Table B1. Descriptive Statistics of Dimension Levels from Datasets 1-3 (Step One)

Variable	Mean (st. err.)	St. Dev.	Skewness (st. err.)	Kurtosis (st. err.)	Min.	Max.
Dataset 1						
IntComply	2.370 (0.012)	1.215	0.000 (0.024)	-0.437 (0.047)	-1.11	5.76
Cel_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
r	-0.005 (0.004)	0.403	0.004 (0.024)	-0.017 (0.047)	-1.62	1.73
Dataset 2						
IntComply	2.375 (0.012)	1.209	0.007 (0.024)	-0.478 (0.047)	-1.29	6.04
Cel_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
r	0.000 (0.004)	0.402	-0.14 (0.024)	-0.039 (0.047)	-1.40	1.47
Dataset 3						
IntComply	3.871 (0.020)	2.115	-0.005 (0.024)	-0.8965 (0.047)	-1.39	9.14

Table B1. Descriptive Statistics of Dimension Levels from Datasets 1-3 (Step One)

Cel_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Cel_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Crt_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Hi	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Med	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
Sev_Low	0.333 (0.005)	0.471	0.707 (0.024)	-1.500 (0.047)	0	1
r	-0.004 (0.004)	0.402	-0.006 (0.024)	0.028 (0.047)	-1.46	1.55
Notes: *N* = 10,800 (400 individuals, 27 vignettes each). Mean of dummy (0, 1) variables indicate frequency of the variable.						

Table B2. Manipulation Effect Sizes of Dimensions from Datasets 1-3 (Step Two)

Dimension	Mean (st. err.)	St. dev.	Skewness (st. err.)	Kurtosis (st. err.)	Min.	Max.
Dataset 1						
IntComply	2.370 (0.012)	1.215	0.000 (0.024)	-0.437 (0.047)	-1.11	5.76
Celerity	0.751 (0.006)	0.615	0.010 (0.024)	-1.500 (0.047)	0.00	1.51
Certainty	0.6257 (0.005)	0.510	-0.005 (0.024)	-1.500 (0.047)	0.00	1.25
Severity	1.005 (0.008)	0.821	0.004 (0.024)	-1.500 (0.047)	0.00	2.01
Dataset 2						
IntComply	2.375 (0.012)	1.209	0.007 (0.024)	-0.478 (0.047)	-1.29	6.04
Celerity	0.750 (0.006)	0.614	0.011 (0.024)	-1.500 (0.047)	0.00	1.51
Certainty	0.622 (0.005)	0.508	0.003 (0.024)	-1.500 (0.047)	0.00	1.25
Severity	0.997 (0.008)	0.815	0.006 (0.024)	-1.500 (0.047)	0.00	2.00
Dataset 3						
IntComply	3.871 (0.020)	2.115	-0.005 (0.024)	-0.896 (0.047)	-1.39	9.14
Celerity	2.258 (0.018)	0.1.84	-0.006 (0.024)	-1.500 (0.047)	0.00	4.51
Certainty	0.628 (0.005)	0.512	-0.008 (0.024)	-1.500 (0.047)	0.00	1.25
Severity	0.993 (0.008)	0.813	0.010 (0.024)	-1.500 (0.047)	0.00	1.99
Notes: manipulation effect sizes are composites of dimension levels displayed in Table B1. *N* = 10,800 (400 individuals, 27 vignettes each).						

Appendix C: Comparison of Type I Sums of Squares for *Cel_Hi* and *Cel_Med*

A mathematical comparison of sequential or Type I sums of squares for *Cel_Hi* and *Cel_Med* explains how larger dimension intervals inflate unstandardized *β* and standardized **B** coefficients of *Celerity* in steps one and two, respectively. We examined two sets of univariate general models for each dataset using IBM SPSS Statistics version 24. The first models (i.e., Table C1: Models C1-1a, C1-2a, and C1-3a) contained the dummy variables for *Crt_Hi*, *Crt_Med*, *Sev_Hi*, and *Sev_Med* as predictors in Datasets 1, 2, and 3, respectively. The second models (i.e., Models C1-1b, C1-2b, and C1-3b) added the dummy variables for *Cel_Hi* and *Cel_Med* to their corresponding first models. We added the *Cel_Hi* and *Cel_Med* dummy variables at the end of SPSS' UNIANOVA command of the first model; therefore, the first models are nested in their corresponding second models.

Results in Table C1 show several marginal effects of increasing the dimension interval of *Celerity* from 0.00 – 1.00 to 0.00 – 3.00 without a corresponding change in the unstandardized *β*. Perhaps the most direct change is the nine-fold increase in the sum of squares for *Cel_Hi* and *Cel_Med* from Model C1-1b (Dataset 1) and Model C1-2b (Dataset 2) to Model C1-3b (Dataset 3). Partitioning the total sum of squares reveals that *Cel_Hi* and *Cel_Med* account for about 76 percent of the total sum of squares in Dataset 3 but only 26 percent in Datasets 1 and 2. These increases also affect the Δ adjusted R^2 statistics, which in Dataset 3 is about three times that of Datasets 1 and 2 (i.e., 0.758 vs. 0.257 and 0.258, respectively). The differences in R^2 and adjusted R^2 values result from the individual sum of squares used to calculate these statistics (e.g., R^2 = 1 – (error sum of squares / total sum of squares)). Figure C1 displays the SPSS code we used to perform the sum of squares analyses of Datasets 1-3.

Table C1. Sum of Squares from Datasets 1-3[a]

	Dataset 1		Dataset 2		Dataset 3	
	Model C1-1a[b]	Model C1-1b[c]	Model C1-2a[b]	Model C1-2b[c]	Model C1-3a[b]	Model C1-3b[c]
Source						
Crt_Hi	2,100.7***	2,100.7***	2,097.5***	2,097.5***	2,112.4***	2,112.4***
Crt_Med	708.1***	708.1***	694.3***	694.3***	717.7***	717.7***
Sev_Hi	5,476.9***	5,476.9***	5,397.8***	5,397.8***	5,385.3***	5,385.3***
Sev_Med	1,811.4***	1,811.4***	1,780.1***	1,780.1***	1,753.0***	1,753.0***
Cel_Hi		3,079.3***		3,075.2***		27,368.7***
Cel_Med		1,003.5***		1,002.2***		9,233.5***
Residual	5,839.1	1,756.4	5,823.9	1,746.5	38,343.0	1,740.8
Total (corrected[d])	15,936.1	15,936.1	15,793.4	15,793.4	48,311.4	48,311.4
Goodness of fit						
R^2	0.634	0.890	0.631	0.889	0.206	0.964
Adjusted R^2	0.633	0.890	0.631	0.889	0.206	0.964
Δ Adjusted R^{2e}		0.257		0.258		0.758

[a] *N* = 10,800 at vignette level; *N* = 400 at individual level. Type I (i.e., "sequential") sum of squares generated with univariate general linear models. Sources in first columns are the dummy variables of the listed parameters.
[b] We entered predictors in Models C1-1a, C1-2a, and C1-3a in the following order: *Crt_Hi*, *Crt_Med*, *Sev_Hi*, and *Sev_Med*.
[c] We entered predictors in Models C1-1b, C1-2b, and C1-3b in the following order: *Crt_Hi*, *Crt_Med*, *Sev_Hi*, *Sev_Med*, *Cel_Hi*, and *Cel_Med*. Thus, Model C1-1a is nested in C1-1b, C1-2a in C1-2b, and C1-3a in C1-2b.
[d] The "corrected total" is the sum of the sums of squares for the main effects and residual term.
[e] Δ Adjusted R^2 is based on difference from corresponding nested model.
*** $p < .001$, ** $p < .01$, * $p < .05$.

```
comment - Sequential Sum of Squares, First Model.
comment - "Build Term(s)" set to "Main effects"; intercept included.
UNIANOVA intComply WITH dumCrtHi dumCrtMed dumSevHi dumSevMed
  /METHOD=SSTYPE(1)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=dumCrtHi dumCrtMed dumSevHi dumSevMed.

comment - Sequential Sum of Squares, Second Model.
comment - "Build Term(s)" set to "Main effects"; intercept not included.
comment – dumCelHi and dumCelMed added at end of UNIANOVA command, so
                the first model is nested within the second.
UNIANOVA intComply WITH dumCrtHi dumCrtMed dumSevHi dumSevMed dumCelHi dumCelMed
  /METHOD=SSTYPE(1)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(.05)
  /DESIGN=dumCrtHi dumCrtMed dumSevHi dumSevMed dumCelHi dumCelMed.
```

Figure C1. SPSS Code for Sum of Squares Analyses of Datasets 1-3

Appendix D: Generation and Analysis of Multilevel Dataset 4

We begin by assuming the reader has a rudimentary knowledge of multilevel modeling. A wide variety of reference texts exists for those who wish to learn more about these techniques. Klein and Kozlowski (2000) is a well-respected text on multilevel theory and techniques. Raudenbush and Bryk (2002) is particularly useful in RCM and the Hierarchical Linear Modeling (HLM) software. Multilevel methods such as hierarchical logit modeling are also available for dichotomous dependent variables (Wong & Mason, 1985).

We constructed illustrative multilevel data following Hofmann and Gavin (1998, pp. 628-633). We generated Dataset 4 from vignettes created from the above 3 x 3 x 3 design under the following specifications. First, three independent variables (i.e., *Celerity*, *Certainty*, and *Severity*) are the vignette-level (level-one) predictors of the dependent variable *IntComply*. We estimated the models using full maximum likelihood. We generated data based on the following equations, which we write in RCM notation. Equation D1 describes the level-one portion of the model:

$$\begin{aligned} IntComply_{iv} = \pi_{0iv} + (\pi_{1iv} * Cel_Low_{iv}) + (\pi_{2iv} * Cel_Med_{iv}) + (\pi_{3iv} * Cel_Hi_{iv}) + \\ (\pi_{4iv} * Crt_Low_{iv}) + (\pi_{5iv} * Crt_Med_{iv}) + (\pi_{6iv} * Crt_Hi_{iv}) + \\ (\pi_{7iv} * Sev_Low_{iv}) + (\pi_{8iv} * Sev_Med_{iv}) + (\pi_{9iv} * Sev_Hi_{iv}) + e_{iv,} \end{aligned} \tag{D1}$$

where *IntComply* represents the dependent variable for individual *i* and vignette *v*; Cel_Low_{iv}, Cel_Med_{iv}, and Cel_Hi_{iv} are dummy (0, 1) variables that represent the presence of statements about low, medium, and high dimension levels, respectively, of the *Celerity* factor; Crt_Low_{iv}, Crt_Med_{iv}, and Crt_Hi_{iv} are dummy (0, 1) variables that represent the presence of statements about low, medium, and high dimension levels, respectively, of the *Certainty* factor; and Sev_Low_{iv} , Sev_Med_{iv}, and Sev_Hi_{iv} are dummy (0, 1) variables that represent the presence of statements about low, medium, and high dimension levels, respectively, of the *Severity* factor; π_{qiv} represents the coefficient of the independent variables of order *q* in the right-hand side of the equation; and e_{iv} represents the vignette-level residuals.

Equations D2 to D8 describe the level-two portion of the model:

$$\pi_0 = \beta_{00} + \beta_{01} RiskAverse + r_0 \tag{D2}$$

$$\pi_1 = \beta_{10} + \beta_{11} RiskAverse + r_1 \tag{D3}$$

$$\pi_2 = \beta_{20} + \beta_{21} RiskAverse + r_2 \tag{D4}$$

$$\pi_3 = \beta_{30} + \beta_{31} RiskAverse + r_3 \tag{D5}$$

$$\pi_4 = \beta_{40} + \beta_{41} RiskAverse + r_4 \tag{D6}$$

$$\pi_5 = \beta_{50} + \beta_{51} RiskAverse + r_5 \tag{D7}$$

$$\pi_6 = \beta_{60} + \beta_{61} RiskAverse + r_{6,} \tag{D8}$$

where $\pi_0 - \pi_6$ represent the level-one intercept and slopes; *e* the level-one residual; $\beta_{00} - \beta_{60}$ the "average values" (Raudenbush & Bryk, 2002, p. 30) of the $\pi_0 - \pi_6$ intercept and slopes, respectively; $\beta_{01} - \beta_{61}$ the slopes between *RiskAverse* and $\pi_0 - \pi_6$, respectively; and $r_1 - r_6$ the level-two residuals.

Table D1 displays the dummy variable values entered into Equation D1 for each vignette used to generate Dataset 4. We set the magnitudes of low-, medium-, and high-level dimension levels (e.g., *Crt_Medmag* represents a medium-level magnitude of factor *Certainty*; *Sev_Himag* represents a high level of magnitude of factor *Severity*) at 0.0, 0.5, and 1.0, respectively. Table D2 displays the parameters we used to generate Dataset 4—which are the same as for Dataset 1, Table 1—for the reader's convenience.

Table D1. Dummy Variable Values by Factor Statement and Vignette: Level One, Step One

Vignette	Statement dummy variables								
	Cel_Hi	***Cel_Med***	***Cel_Low***	***Crt_Hi***	***Crt_Med***	***Crt_Low***	***Sev_Hi***	***Sev_Med***	***Sev_Low***
1	0	0	1	0	0	1	0	0	1
2	0	0	1	0	0	1	0	1	0
3	0	0	1	0	0	1	1	0	0
4	0	0	1	0	1	0	0	0	1
5	0	0	1	0	1	0	0	1	0
6	0	0	1	0	1	0	1	0	0
7	0	0	1	1	0	0	0	0	1
8	0	0	1	1	0	0	0	1	0
9	0	0	1	1	0	0	1	0	0
10	0	1	0	0	0	1	0	0	1
11	0	1	0	0	0	1	0	1	0
12	0	1	0	0	0	1	1	0	0
13	0	1	0	0	1	0	0	0	1
14	0	1	0	0	1	0	0	1	0
15	0	1	0	0	1	0	1	0	0
16	0	1	0	1	0	0	0	0	1
17	0	1	0	1	0	0	0	1	0
18	0	1	0	1	0	0	1	0	0
19	1	0	0	0	0	1	0	0	1
20	1	0	0	0	0	1	0	1	0
21	1	0	0	0	0	1	1	0	0
22	1	0	0	0	1	0	0	0	1
23	1	0	0	0	1	0	0	1	0
24	1	0	0	0	1	0	1	0	0
25	1	0	0	1	0	0	0	0	1
26	1	0	0	1	0	0	0	1	0
27	1	0	0	1	0	0	1	0	0

Table D2. Parameters Used to Generate Dataset 4[a]

Vignette-level parameter	Datasets 1 & 4	Dataset 2	Dataset 3
IntComply-intercept (β_0)	0.00	0.00	0.00
β_1 (*Celerity* coefficient)	1.50	0.50	1.50
Cel_Low / *Cel_Med* / *Cel_Hi* (dimension-level) effects	0.00 / 0.50 / 1.00	0.00 / 1.50 / 3.00	0.00 / 1.50 / 3.00
β_2 (*Certainty* coefficient)	1.25	1.25	1.25
Crt_Low / *Crt_Med* / *Crt_Hi* (dimension-level) effect size	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00
β_3 (*Severity* coefficient)	2.00	2.00	2.00
Sev_Low / *Sev_Med* / *Sev_Hi* (dimension-level) effect size	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00	0.00 / 0.50 / 1.00
Residual (e_{ti})	0.00 (0.40)	0.00 (0.40)	0.00 (0.40)

[a] We used Dataset 1 parameters to generate Dataset 4. We provide the parameters for Datasets 2 and 3 as a reference. We include standard deviations in parentheses where applicable (e.g., for residuals). *N* = 10,800 (400 individuals, 27 vignettes each). Coefficients and dimension-level effect sizes for variables *Certainty* and *Severity* (e.g., β_2, *Crt_Low*, and *Crt_Hi*; β_3, *Sev_Low*, and *Sev_Hi*) remain constant across all three datasets.

Table D3. Level-two Intercepts β_{00} – β_{60} (for Level-one Coefficients π_1 – π_6): Step One

Model	Level-two coefficient β_{0j} (intercept of cross-level effect of *RiskAverse*)			
	Intercept[a]	*Celerity*[a]	*Certainty*	*Severity*
4-1	0.00	1.50	1.25	2.00
	Level-two coefficient β_{1j} (coefficient of cross-level effect of *RiskAverse*)			
4-1	0.25	0.4	0.00	0.00

[a] Values listed in this column are for the constants β_{i0} in the level-two equations $\pi_{ij} = \beta_{i0} + \beta_{i1}RiskAverse + u_{ij}$. The value of π_{ij} will vary in multilevel models when *RiskAverse* is added as a level-two moderator.

Figure D1 provides the R code we used to generate Dataset 4 for Tables 4, 6, and 8.

```
# R code to generate Dataset 4 for Tables 4, 6, and 8.
# Datasets will be stored in the working directory in a new folder, “FSM_Datasets”.
#
# Constants used throughout this file.
# Dimension-level effect sizes. These are equal to Dataset 1, Table 1.
celHi <- 1.0
celMed <- 0.5
celLow <- 0.0
crtHi <- 1.0
crtMed <- 0.5
crtLow <- 0.0
sevHi <- 1.0
sevMed <- 0.5
sevLow <- 0.0
# Beta coefficients. These are equal to Dataset 1, Table 1.
celCoeff <- 1.50
crtCoeff <- 1.25
sevCoeff <- 2.00
# Residual standard deviations. Ensure the e_sd is the same as Dataset 1.
e_sd <- 0.4   # Level-1
r0_sd <- 0.2  # Level-2 residual for level-1 intercept.
r1_sd <- 0.2  # Level-2 residual for level-2 coefficient for celHi
r2_sd <- 0.2  # Level-2 residual for level-2 coefficient for celMed
r3_sd <- 0.2  # Level-2 residual for level-2 coefficient for celLow
# Level-2 values.
```

```
riskAverseMean <- 3.0
riskAverse_sd <- 1.25
fixedGamma01 <- 0.25
fixedGamma11 <- 0.4

# Generate VignID numbers.
vignID <- seq(from = 1001, to = 11800, by = 1)
#
# Generate Count numbers.
for(i in 1:400){
  if(i <= 1) {
    count <- seq(from = 1, to = 27, by = 1)
  } else {
    count <- c(count, seq(from = 1, to = 27, by = 1))
  }
}
#
# Generate individual (subject) ID numbers.
indivID_400 <- seq(from = 101, to = 500, by = 1)
indivID <- c(101, 101)
for(i in 3:27) {
  indivID <- c(indivID, 101)
}
for(i in 102:500) {
  for(j in 1:27) {
    indivID <- c(indivID, i)
  }
}
#
# Generate residuals. Certainty and Severity dummy variable coefficients are
# fixed, so residuals are not needed for those four variables.
e <- rnorm(10800,mean=0,sd=e_sd)
#
# Generate level-two RiskAverse and residual values.
riskAverse_temp <- rnorm(1,mean=riskAverseMean,sd=riskAverse_sd)
r0_temp <- rnorm(1,mean=0,sd=r0_sd)
r1_temp <- rnorm(1,mean=0,sd=r1_sd)
r2_temp <- rnorm(1,mean=0,sd=r2_sd)
r3_temp <- rnorm(1,mean=0,sd=r3_sd)
riskAverse_400 <- riskAverse_temp  # This generates a list of the 400
r0_400 <- r0_temp              #individual riskAverse values by individual.
r1_400 <- r1_temp
r2_400 <- r2_temp
r3_400 <- r3_temp
for(i in 2:400) {
  # Generate a level-2 vectors with 400 items
  riskAverse_temp <- rnorm(1,mean=riskAverseMean,sd=riskAverse_sd)
  r0_temp <- rnorm(1, mean=0, sd=r0_sd)
  r1_temp <- rnorm(1, mean=0, sd=r1_sd)
  r2_temp <- rnorm(1, mean=0, sd=r2_sd)
  r3_temp <- rnorm(1, mean=0, sd=r3_sd)
  riskAverse_400 <- rbind(riskAverse_400, riskAverse_temp) # This generates a list of the 400
  r0_400 <- rbind(r0_400, r0_temp)                #individual level-2 values by
  r1_400 <- rbind(r1_400, r1_temp)                # individual.
  r2_400 <- rbind(r2_400, r2_temp)
  r3_400 <- rbind(r3_400, r3_temp)
}
#
```

```
# Vignette-level independent (dummy) variables
for(i in 1:3) {
  if(i == 1) {
    percdCel <- c(0, 0, 1)
  } else if(i == 2) {
    percdCel <- c(0, 1, 0)
  } else {
    percdCel <- c(1, 0, 0)
  }
  for(j in 1:3) {
    if(j == 1) {
      percdCrt <- c(0, 0, 1)
    } else if(j == 2) {
      percdCrt <- c(0, 1, 0)
    } else {
      percdCrt <- c(1, 0, 0)
    }
    for(k in 1:3) {
      if(k == 1) {
        percdSev <- c(0, 0, 1)
      } else if(k == 2) {
        percdSev <- c(0, 1, 0)
      } else {
        percdSev <- c(1, 0, 0)
      }
      temp <- c(percdCel, percdCrt, percdSev)
      if(i == 1 & j == 1 & k == 1) {
        vignetteComp <- temp
      } else {
        vignetteComp <- rbind(vignetteComp, temp)
      }
     }
  }
}
#
# Moderating and direct effects on the dependent variable, Intention to Comply (intComply).
lvl1_count <- 0    # Level-1 counter variable for following for loops, 1 to 10800
riskAverseCheck_10800 <- 0
eCheck_10800 <- 0
r0Check_10800 <- 0
r1Check_10800 <- 0
r2Check_10800 <- 0
r3Check_10800 <- 0
for(i in 1:400) {
  for(j in 1:27) {
    lvl1_count <- lvl1_count + 1
    tp0 <- (fixedGamma01 * riskAverse_400[i]) + r0_400[i]
    tp1 <- (vignetteComp[j,1] * celHi) * (celCoeff + (fixedGamma11 * riskAverse_400[i]) + r1_400[i])
    tp2 <- (vignetteComp[j,2] * celMed) * (celCoeff + (fixedGamma11 * riskAverse_400[i]) + r2_400[i])
    tp3 <- (vignetteComp[j,3] * celLow) * (celCoeff + (fixedGamma11 * riskAverse_400[i]) + r3_400[i])
    tp4 <- (vignetteComp[j,4] * crtHi * crtCoeff)
    tp5 <- (vignetteComp[j,5] * crtMed * crtCoeff)
    tp6 <- (vignetteComp[j,6] * crtLow * crtCoeff)
    tp7 <- (vignetteComp[j,7] * sevHi * sevCoeff)
    tp8 <- (vignetteComp[j,8] * sevMed * sevCoeff)
    tp9 <- (vignetteComp[j,9] * sevLow * sevCoeff)
    temp <- tp0 + tp1 + tp2 + tp3 + tp4 + tp5 + tp6 + tp7 + tp8 + tp9 + e[lvl1_count]
    if(i == 1 & j == 1) {
```

```
      intComply <- temp
      # Initialize check variables to be printed in level-1 output.
      riskAverseCheck_10800 <- riskAverse_400[i]
      eCheck_10800 <- e[lvl1_count]
      r0Check_10800 <- r0_400[i]
      r1Check_10800 <- r1_400[i]
      r2Check_10800 <- r2_400[i]
      r3Check_10800 <- r3_400[i]
    } else {
      intComply <- rbind(intComply, temp)
      # Update check variables to be printed in level-1 output.
      riskAverseCheck_10800 <- rbind(riskAverseCheck_10800, riskAverse_400[i])
      eCheck_10800 <- rbind(eCheck_10800, e[lvl1_count])
      r0Check_10800 <- rbind(r0Check_10800, r0_400[i])
      r1Check_10800 <- rbind(r1Check_10800, r1_400[i])
      r2Check_10800 <- rbind(r2Check_10800, r2_400[i])
      r3Check_10800 <- rbind(r3Check_10800, r3_400[i])
    }

  }
}
#
# Create level-1 csv file.
# Add "FSM_Datasets" folder to working directory if necessary.
if(file.exists(file.path(getwd(),"FSM_Datasets"))) {
    # If the FSM_Datasets folder exists, do nothing.
} else {
    dir.create(file.path(getwd(),"FSM_Datasets"))
}
fileContents <- cbind(vignID, count, indivID, intComply, eCheck_10800, riskAverseCheck_10800,
                      r0Check_10800, r1Check_10800, r2Check_10800, r3Check_10800)

# Copy the celHi-celMed...sevLow matrix for each individual.
vignetteComp400 <- vignetteComp
for(i in 2:400) {
  vignetteComp400 <- rbind(vignetteComp400, vignetteComp)
}
fileContents <- cbind(fileContents, vignetteComp400)

# Name columns.
# RiskAverse and residuals (both levels one and two) are included in output for verification
# purposes
colnames(fileContents) <- c("VignID", "Count", "IndivID", "IntComply",
                            "e_lvl1Check", "RiskAverse_lvl1Ccheck",
                            "r0_lvl1Check","r1_lvl1Check","r2_lvl1Check","r3_lvl1Check",
                            "Cel_Hi", "Cel_Med", "Cel_Low",
                            "Crt_Hi", "Crt_Med", "Crt_Low",
                            "Sev_Hi", "Sev_Med", "Sev_Low")
write.csv(fileContents, file="FSM_Datasets/R_FigureD1_Dataset4_Level1.csv",
          row.names = FALSE)

# Create level-2 csv file.
fileContents2 <- cbind(indivID_400, riskAverse_400, r0_400, r1_400, r2_400, r3_400)
colnames(fileContents2) <- c("indivID", "RiskAverse", "r0", "r1", "r2", "r3")
# Level-2 residuals are stored in "R_FigureD1_Dataset4_Level1.csv".
write.csv(fileContents2, file="FSM_Datasets/R_FigureD1_Dataset4_Level2.csv",
          row.names = FALSE)
```

Figure D1. R Code to Generate Dataset 4

Table D4. Descriptive Statistics for Dataset 4

Parameters used to generate dataset [a]			
Level-one parameter	**Level**	**Level-two parameter**	**Mean**
π_{0i} (intercept)	1	β_{00} (intercept)	0.00
		β_{01} (*RiskAverse* coefficient)	0.25
		u_0 (residual)	0.00 (0.10)
π_{1i} (*Cel_Hi* coefficient)[b]	1	β_{10} (intercept)	1.50
		β_{11} (*RiskAverse* coefficient)	0.40
		u_1 (residual)	0.00 (0.10)
π_{2i} (*Cel_Med* coefficient)[b]	1	β_{20} (intercept)	0.75
		β_{21} (*RiskAverse* coefficient)	0.00
		u_2 (residual)	0.00 (0.10)
π_{3i} (*Cel_Low* coefficient) [b]	1	β_{30} (intercept)	0.00
		β_{31} (*RiskAverse* coefficient)	0.00
		u_3 (residual)	0.00 (0.10)
π_{4i} (*Crt_Hi* coefficient)[b]	1	β_{40} (intercept)	1.25
		β_{41} (*RiskAverse* coefficient)	0.40
		u_4 (residual)	0.00 (0.10)
π_{5i} (*Crt_Med* coefficient)[b]	1	β_{50} (intercept)	0.75
		β_{51} (*RiskAverse* coefficient)	0.00
		u_5 (residual)	0.00 (0.10)
π_{6i} (*Crt_Low* coefficient) [b]	1	β_{60} (intercept)	0.00
		β_{61} (*RiskAverse* coefficient)	0.00
		u_6 (residual)	0.00 (0.10)
π_{7i} (*Sev_Hi* coefficient)[b]	1	β_{70} (intercept)	2.00
		β_{71} (*RiskAverse* coefficient)	0.40
		u_7 (residual)	0.00 (0.10)
π_{8i} (*Sev_Med* coefficient)[b]	1	β_{80} (intercept)	1.00
		β_{81} (*RiskAverse* coefficient)	0.40
		u_8 (residual)	0.00 (0.10)
π_{9i} (*Sev_Low* coefficient)[b]	1	β_{90} (intercept)	1.50
		β_{91} (*RiskAverse* coefficient)	0.40
		u_9 (residual)	0.00 (0.10)
r_{ti} (residual)	1	N/A	0.00 (0.40)
RiskAverse	2	N/A	3.00 (1.25)

Descriptive statistics of generated dataset									
Variable	**Mean**	**St. err. of mean**	**St. dev.**	**Skewness**	**St. err. of skewness**	**Kurtosis**	**St. err. of kurtosis**	**Min.**	**Max.**
Level one[c]									
IntComply	3.674	0.016	1.643	0.177	0.024	-0.363	0.047	-0.81	9.24
Cel_Hi	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Cel_Med	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Cel_Low	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Crt_Hi	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0

Table D4. Descriptive Statistics for Dataset 4

Crt_Med	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Crt_Low	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Sev_Hi	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Sev_Med	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
Sev_Low	0.33	0.005	0.471	0.707	0.024	-1.500	0.047	0.0	1.0
e	0.00	0.004	0.397	.001	0.024	-0.034	0.047	-1.49	1.46
Level two[c]									
RiskAverse	2.925	0.062	1.246	-0.110	0.122	0.417	0.243	-1.35	7.56
r_0	-0.012	0.010	0.197	-0.101	0.122	0.084	0.243	-0.67	0.057
r_1	-0.005	0.011	0.216	0.165	0.122	-0.016	0.243	-0.541	0.684
r_2	-0.015	0.010	0.200	0.105	0.122	-0.332	0.243	-0.505	0.628
r_3	0.007	0.010	0.204	0.236	0.122	0.030	0.243	-0.487	0.617

[a] We include standard deviations in parentheses where applicable (e.g., for residuals). Subscripts will not match models in the text when low-dimension dummy variables are omitted to preclude linear dependences.
[b] We generated high, medium, and low coefficients by multiplying this base factor by 1.00, 0.50, and 0.00, respectively. For example, the *Celerity* coefficient for the high dummy variable = 1.50 * 1.00 = 1.50, for the medium dummy variable = 1.50 * 0.50 = 0.75, and for the low dummy variable = 1.50 * 0.00 = 0.00. These coefficients would then be multiplied by the value of the corresponding dummy variable. If a vignette contained *Cel_Med* (and not *Cel_Hi* or *Cel_Low*), the *Celerity* effect would be (*Cel_Hi* * 1.50) + (*Cel_Med* * 0.75) + (*Cel_Low* * 0.00) = (0 * 1.50) + (1 * 0.75) + (0 * 0.00) = 0.75.
[c] *N* = 10,800 (400 individuals, 27 vignettes each) for all statistics except *RiskAverse*, where *N* = 400.

Figure D2 displays the R code for CPE step two in Tables 5, 7, and 9 (Dataset 4).

```
# R code to generate step two (CPE) of Dataset 4 for Tables 5, 7, and 9.
# Do not use for Table 3!
# Both tables read the following file, but beta coefficients diff in lines 10 and 11.

# Enter the correct modelNum below.
modelNum <- 91 # Enter "5", "7", "91" (Model 9-1), or "92" (Model 9-2).
#
data <- read.csv(file = "FSM_Datasets/R_FigureD1_Dataset4_Level1.csv")
#
# Replace "n.nnn" with beta coefficients from step one.
#
Celerity_High <- n.nnn
Celerity_Med <- n.nnn
Certainty_High <- n.nnn
Certainty_Med <- n.nnn
Severity_High <- n.nnn
Severity_Med <- n.nnn
#
# Generate aggregate data variables
Celerity <- (data[11] * Celerity_High) + (data[12] * Celerity_Med)
Certainty <- (data[14] * Certainty_High) + (data[15] * Certainty_Med)
Severity <- (data[17] * Severity_High) + (data[18] * Severity_Med)
#
# Create csv file that mimics Table10Data. Dummy variables and RiskAverse included for
# references purposes
fileContents <- cbind(data[,1:4], data[6],data[,11:19], Celerity, Certainty, Severity)
colnames(fileContents) <- c("vignID", "count", "indivID", "IntComply", "RiskAverse",
                            "dumCelHi", "dumCelMed", "dumCelLow",
                            "dumCrtHi", "dumCrtMed", "dumCrtLow",
                            "dumSevHi", "dumSevMed", "dumSevLow",
                            "Celerity", "Certainty", "Severity")
# Create csv output file.
```

```
if (modelNum == 5) {
   write.csv(fileContents, file="FSM_Datasets/R_Table_5_Step2CPE.csv",
         row.names = FALSE)
} else if (modelNum == 7) {
   write.csv(fileContents, file="FSM_Datasets/R_Table_7_Step2CPE.csv",
         row.names = FALSE)
} else if (modelNum == 91) {
  write.csv(fileContents, file="FSM_Datasets/R_Model_9-1_Step2CPE.csv",
         row.names = FALSE)
} else {
  write.csv(fileContents, file="FSM_Datasets/R_Model_9-2_Step2CPE.csv",
         row.names = FALSE)
}
```

Figure D2. R Code to Generate Step Two (CPE) Results for Tables 5, 7, and 9

Appendix E: HLM for Windows Code for Tables 6-9

Figures E1 and E2 show the HLM for Windows code for Models 6-4, 8-3, and 8-4. Note that the main difference between the two figures is that the level-two variable *RiskAverse* is uncentered in Models 6-4 and 8-3 (as represented by plain font) but grand-mean centered in Model 8-4 (as represented by bold-italic font). All variables in the level-one model in Figure E1 are dummy variables. The level-one coefficients π_3 – π_6 are fixed across groups in both models. Also note that variable names in HLM are capitalized and clipped to an eight-character length (e.g., *IntComply* to INTCOMPL and *RiskAverse* to RISKAVER). Figure E3 displays the HLM for Windows code for Table 7. We ran all HLM models using full maximum likelihood estimation.

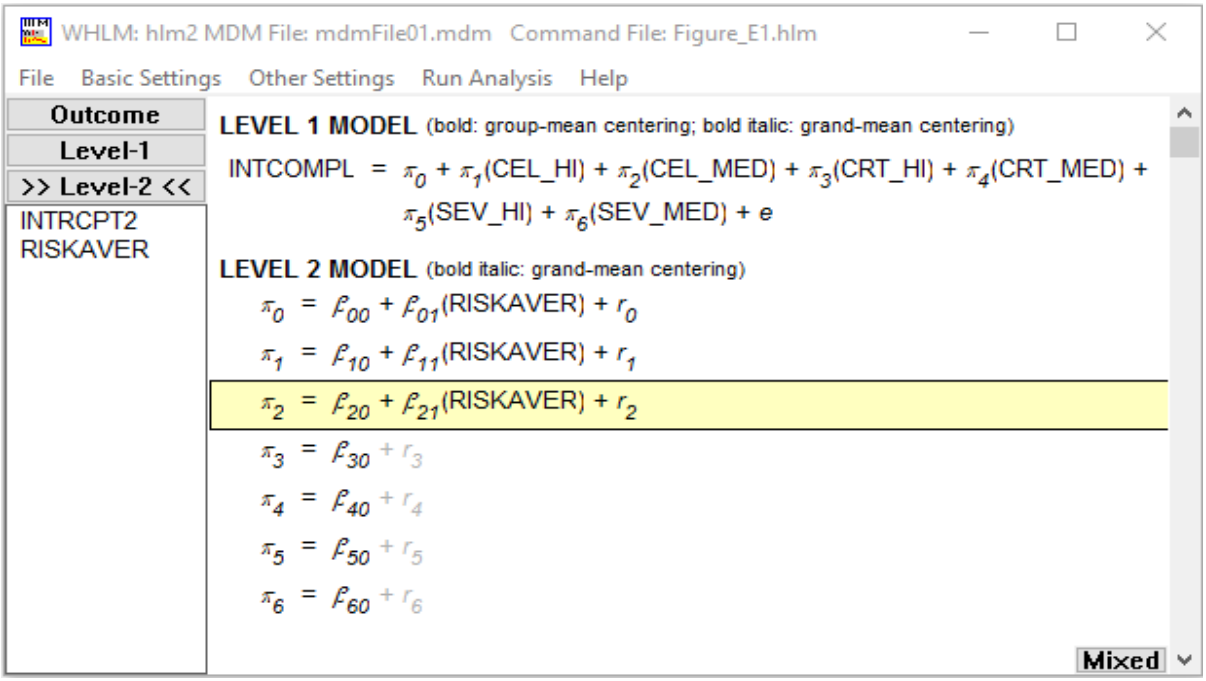

Figure E2. HLM for Windows Code for Tables 6-4 and 8-3

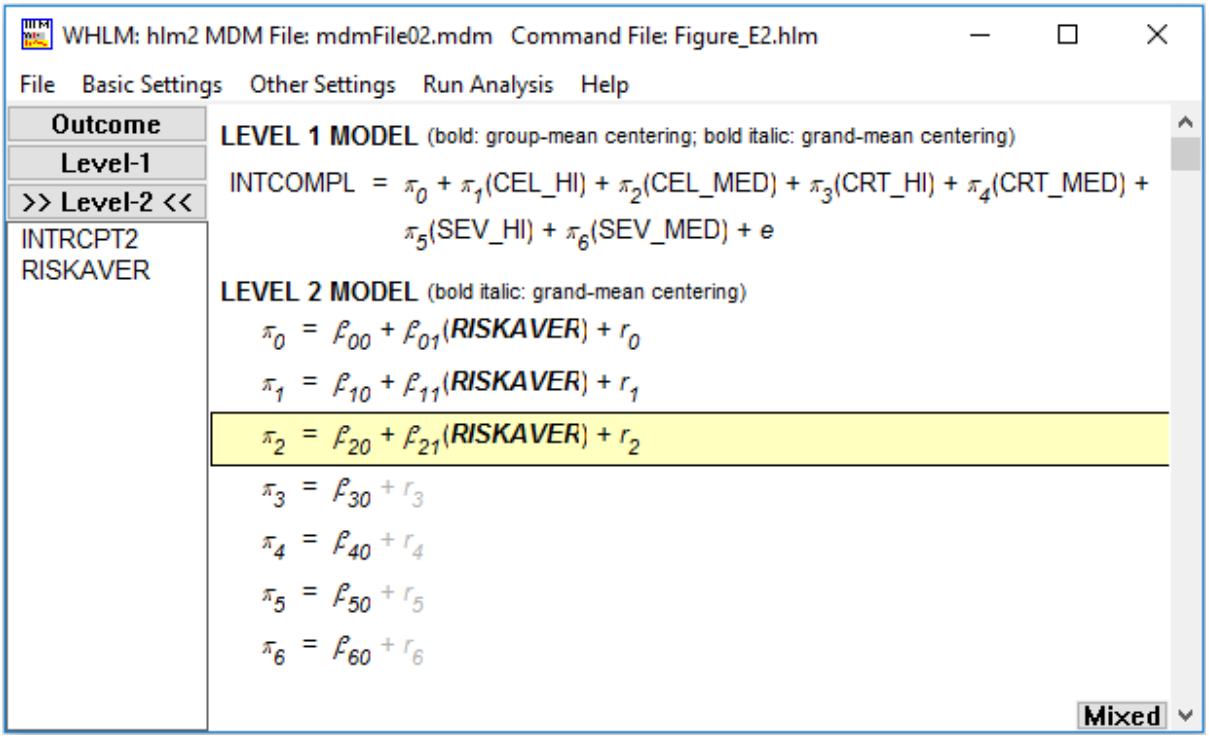

Figure E2. HLM Code for Model 8-4

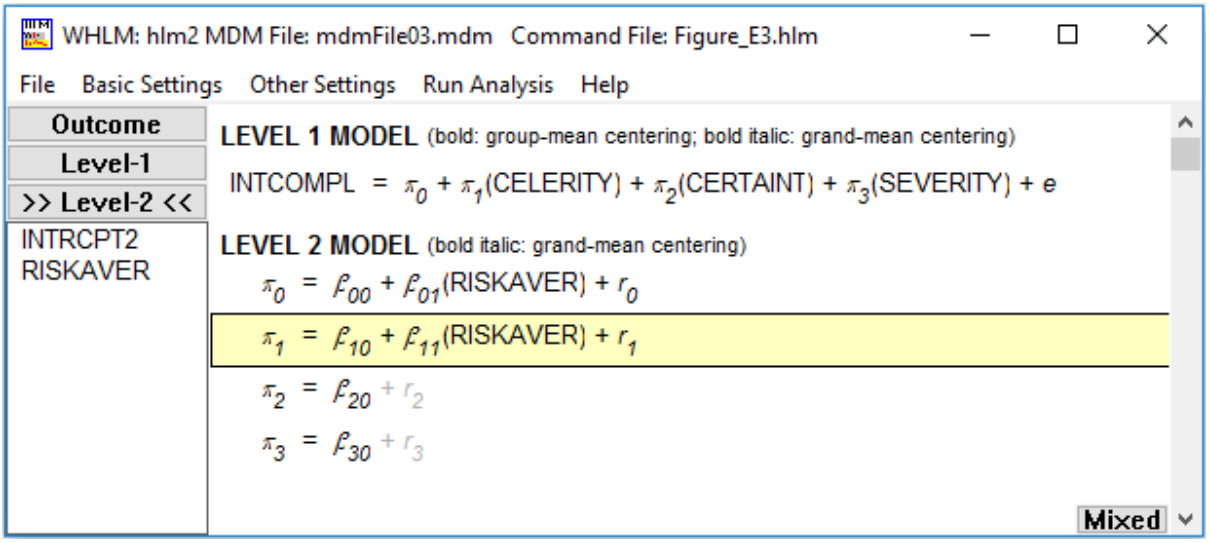

Figure E3. HLM for Windows Code for Table 7

Appendix F: Some Effects of Inconsistent Centering Across Steps 1 and 2

It is useful to examine how inconsistent centering affects CPE step two (Table F1). Note that the procedures we used to calculate Table F1 are for pedagogical purposes only—we do *not* consider them to be best practices (or even fairly reasonable ones). Essentially, Table F1 mirrors Table 9 except for the inconsistent centering of the level-two variable *RiskAverse*. Models 9-1 and 9-2 each match the types of centering in steps one and two (i.e., Model 9-1 uses uncentered *RiskAverse* and Model 9-2 uses $RiskAverse_{GMC}$). Table F1, on the other hand, demonstrates how incorrect results arise when centering mismatches occur between steps one and two.

Table F1. Effects of Inconsistent Centering on HLM Analyses of Factorial Surveys: Step Two (Dataset 4)[a]

Model	True value	F1-1			F1-2		
Level-one (*Celerity*, *Certainty*, *Severity*)		Uncentered					
CPE process		Step one inputs[b]	Step two results		Step one inputs[b]	Step two results	
Level-two *RiskAverse* centering		Uncentered	GMC		GMC	Uncentered	
Coefficient		Model 8-3 β	β	**B**[c]	Model 8-4 β	β	**B**[c]
Level-one intercept and independent variables (coefficients) $\pi_0 - \pi_3$							
π_0 (β_{00}, level-two intercept for *IntComply*-intercept π_0)	0.00	N/A	0.694*** (0.013)		N/A	-0.020 (0.029)	
π_1 (β_{10}, level-two intercept for *Celerity* coefficient π_1)	*Cel_Hi*: 1.50 *Cel_Med*: 0.75	*Cel_Hi*: 1.497 *Cel_Med*: 0.753	1.788*** (0.010)	0.665***	*Cel_Hi*: 2.677 *Cel_Med*: 1.324	0.559*** (0.015)	0.372***
π_2 (β_{20}, level-two intercept for *Certainty* coefficient π_2)[e]	*Crt_Hi*: 1.25 *Crt_Med*: 0.625	*Crt_Hi*: 1.259 *Crt_Med*: 0.646	1.000*** (0.002)	0.311***	*Crt_Hi*: 1.259 *Crt_Med*: 0.646	1.000*** (0.007)	0.311***
π_3 (β_{30}, level-two intercept for *Severity* coefficient π_3)[e]	*Cel_Hi*: 2.00 *Cel_Med*: 1.00	*Sev_Hi*: 2.008 *Sev_Med*: 1.003	1.000*** (0.001)	0.500***	*Sev_Hi*: 2.008 *Sev_Med*: 1.003	1.000*** (0.004)	0.500***
Direct and moderating effects of level-two *RiskAverse* on level-one coefficients $\pi_0 - \pi_3$							
β_{01} (π_0, *IntComply*-intercept)			0.246*** (0.009)			0.247*** (0.009)	
β_{11} (π_1, *Celerity*)			0.270*** (0.008)			0.151*** (0.004)	
Goodness of fit							
Deviance	NA		12,142.8			12,132.3	
Parameters	NA		17			17	

[a] N = 10,800 at vignette level; N = 400 at individual level. Unstandardized coefficient estimates and robust standard errors (in parentheses) reported.
[b] These step one β values are used as inputs to step two CPE models.
[c] **B** values calculated using Equation 1. s_x and s_y values are taken from the HLM for Windows' MDM template file.
*** $p < .001$, ** $p < .01$, * $p < .05$.

Perhaps the most notable change in Table F1 is that the step two β coefficients of *Celerity* in Models F1-1 and F1-2 do not equal or even approximate 1.000. Instead, the values widely differ: 1.788 in Model F1-1

and 0.559 for Model F1-2. These results by themselves suggest something is amiss, though the differences in how we modeled *RiskAverse* can explain them. Using Equation 3 and assuming $r_1 = 0$, we can calculate the β_{10} value for Model F1-1 (step two) when *RiskAverse* equals $\mu_{RiskAverse}$ rather than 0.00:

$$\pi_1 = \beta_{10} + \beta_{11} RiskAverse + r_1 \quad \text{(F1)}$$

$$1.788 = \beta_{10} + 0.270(2.925) + (0) \quad \text{(F2)}$$

$$\beta_{10} = 1.788 - 0.790 + (0) \quad \text{(F3)}$$

$$\beta_{10} = 0.998 \quad \text{(F4)}$$

Calculating the β_{10} value in Model F1-2, however, is not as straightforward because the step one inputs (i.e., β values) in Model F1-2 are based on commingled effects of both *Celerity* and *RiskAverse*. As in our explanation of the *Cel_Hi* and *Cel_Med* coefficients in Model 8-4, the β_{10} value for *Celerity* in step two of Model F1-2 is based on $RiskAverse_{GMC} = 0.00$ (i.e., $\mu_{RiskAverse} = 2.925$, where $\mu_{RiskAverse}$ equals the grand mean of *RiskAverse*). Thus, factoring out *RiskAverse*'s influence—and by extension, that of $RiskAverse_{GMC}$ as well—requires an equivalent value of $RiskAverse_{GMC}$ for *RiskAverse* = 0.00. Since *RiskAverse* = 2.925 is equivalent to $RiskAverse_{GMC} = 0.00$, then subtracting 2.925 (i.e., the grand mean value of *RiskAverse*) from both sides of each equation produces a *RiskAverse* value of 0.00 and a $RiskAverse_{GMC}$ value of -2.925. Using this latter value of $RiskAverse_{GMC}$, we can now calculate the value of β_{10}:

$$\pi_1 = \beta_{10} + \beta_{11} RiskAverse_{GMC} + r_1 \quad \text{(F5)}$$

$$0.559 = \beta_{10} + 0.151(-2.925) + (0) \quad \text{(F6)}$$

$$\beta_{10} = 0.559 - (-0.442) \quad \text{(F7)}$$

$$\beta_{10} = 1.001 \quad \text{(F8)}$$

Equations F1-F4 and F5-F6 demonstrate that, while both Models F1-1 and F1-2, respectively, are ultimately based on a β_{10} intercept (i.e., the "average value" of π_1) of 1.000, the mismatched centering between steps one and two obscures that foundation and creates confusion. Using these incorrect β coefficients will, in turn, create further mistakes in the calculation of standardized **B** coefficients and their respective rankings, which we now explore. Based on Equation 1, the value for **B** in Model F1-1 is **B** = (β * s_x)/s_y = (1.788 * 0.61)/1.64 = 0.665; for Model F1-2, **B** = (0.559 * 1.09)/1.64 = 0.372. Thus, *Celerity* is ranked *higher* than *Severity* in Model F1-1 but *lower* than *Severity* in Model F1-2 (i.e., the opposite rankings in Models 9-1 and 9-2). The essence of the problem is the inconsistent modeling of the targeted subject; that is, in one step the researcher models *Celerity*, *Certainty*, and *Severity* after controlling for *RiskAverse* (i.e., on a subject exhibiting *RiskAverse* = 0.00), while, in the other step, the researcher models *Celerity*, *Certainty*, and *Severity* based on a subject who exhibits an "average" level of *RiskAverse* of 2.925.

About the Authors

Robert F. Otondo is an Associate Professor of Information Systems at Mississippi State University. He received his PhD in Computer Information Systems at Arizona State University. His research interests center on perceptions and uses of emerging technologies. His research has been funded by the National Science Foundation, the Office of Naval Research, the Robert Wood Johnson Foundation, and the FedEx Center for Supply Chain Management at The University of Memphis. His research has been published in *MIS Quarterly*, *Journal of Applied Psychology, European Journal of Information Systems*, *Information & Management*, *Communications of the Association for Information Systems*, *Personnel Psychology*, *Production and Operations Management*, *Decision Support Systems*, and *Human Relations*.

Robert E. Crossler is an Assistant Professor of Information Systems in the Carson College of Business at Washington State University. He received his PhD from the Department of Accounting and Information Systems at Virginia Tech. His research focuses on the factors that affect the security and privacy decisions individuals make. He has published in leading MIS journals, including *MIS Quarterly*, *Journal of the AIS*, *Information Systems Journal*, *Decision Support Systems*, and *The DATA BASE for Advances in Information Systems*. He received the 2013 INFORMS Information Systems Society (ISS) Design Science Award for his information privacy work, his paper in *The DATA BASE for Advances in Information Systems* was recognized as the journal's best paper in 2014, and he received the *Journal of Information Systems* inaugural Best Paper Award in 2017.

Merrill Warkentin is the James J. Rouse Professor of Information Systems in the College of Business at Mississippi State University. His research, primarily on the impacts of organizational, contextual, and dispositional influences on individual behaviors in the context of information security and privacy, has appeared in *MIS Quarterly*, *Journal of MIS*, *Journal of the AIS*, *Decision Sciences*, *European Journal of Information Systems*, *Information & Management*, *Decision Support Systems*, *Information Systems Journal*, and others. He is the author or editor of seven books, and has authored about 300 published manuscripts, including 75 peer-reviewed journal articles. He was Program Co-Chair of AMCIS2016 and he serves or has served in editorial roles at *MIS Quarterly*, *Information Systems Research*, *Decision Sciences*, *European Journal of Information Systems*, *Information & Management*, and *AIS Transactions on Replication Research*. His work has been funded by NATO, NSF, NSA, DoD, Homeland Security, IBM, and others.

Copyright © 2018 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.