Restricted Variance Interaction Effects: What they are and why they are your friends

Jose M. Cortina

Virginia Commonwealth University

Tine Koehler

University of Melbourne

Kathleen R. Keeler

Virginia Commonwealth University

Bo B. Nielsen

University of Sydney

Acknowledgements: The authors wish to express their gratitude to Ernest O’Boyle and two anonymous reviewers for their insightful comments. Work on the paper was supported by the University of Melbourne’s Faculty of Business and Economics and by the University of Sydney Business School.

Corresponding Author:

Jose Cortina
School of Business
Virginia Commonwealth University
Richmond, VA, 23284
jcortina@vcu.edu
8048272022

ABSTRACT

Although interaction hypotheses are increasingly common in our field, many recent papers have pointed out that authors often have difficulty justifying interaction hypotheses. The purpose of this paper is to describe a particular type of interaction, the restricted variance (RV) interaction. The essence of the RV interaction is that, as the value of one variable in a system changes, certain values of another variable in the system become less plausible, thus restricting its variance. This, in turn, influences relationships between that variable and other variables. These types of interactions are quite common, even if they are not recognized as RV interactions, and they exist at every level of analysis. The advantage of the RV interaction is that, compared to other interaction types, it is relatively simple to justify. The different forms of RV interaction do, however, contain complexities of which a researcher must be aware. This paper explains and illustrates the forms that RV interactions can take and the implications, often counterintuitive, of many of the forms. It also describes how one should go about testing them. Our intention for this paper is to help researchers strengthen and focus their interaction arguments.

**RESTRICTED VARIANCE INTERACTION EFFECTS:
WHAT THEY ARE AND WHY THEY ARE YOUR FRIENDS.**

Interaction models are increasingly common in the organizational sciences. Indeed, it would appear that this trend has been in place for some time (cf. Aguinis, 1995; Cortina, 2003; Cortina, 1993). Even a cursory review of any of our top journals will show that it is rare to come across an empirical paper (or a theoretical one) that doesn't include an interaction hypothesis of some sort (Aguinis, Edwards, & Bradley, 2017).

With this popularity has come concern over what appears to be confusion regarding the conceptualization and testing of interaction models. For example, Andersson, Cuervo-Cazurra, and Nielsen (2014) point out that authors sometimes justify interactions by arguing for the effect of the moderator on the dependent variable even though such an argument is not only insufficient, it is unnecessary for many types of interactions. Holland, Shore, and Cortina (2017) point out that authors sometimes justify what amounts to a particular sign for the product term weight in a moderated regression equation, without considering the fact that the pattern of an interaction is determined not only by the product term weight but also by the weights attached to the variables that make up the product. As another example, in cross-level research, authors sometimes argue that because a putative moderator variable is likely to vary across units, within-unit relationships will vary across levels of the moderator (e.g., Chowdhury & Endres, 2010). Lack of clarity regarding the nature of interaction effects may lead us to pursue interaction effects that don't really have any basis, which would explain why effect sizes for interactions are often relatively small (Aguinis, Beaty, Boik, & Pierce, 2005; Murphy & Russell, 2017).

In a recent editorial, Cortina, Köhler, and Nielsen (2015) argued that a particular category of interaction hypothesis – the restricted variance interaction – is quite common in international

business (IB) research, but most IB researchers don't recognize them. The essence of the restricted variance (RV) interaction is that, as the value of one variable in a system changes, certain values of another variable in the system become less plausible, thus restricting its variance. This, in turn, influences relationships between that variable and other variables. Thus, a variable Z acts as a moderator of the relationship between X and Y because the variance of X or Y is compressed at certain levels of Z, and this compression affects the X-Y relationship. As we argue below, just as these sorts of interactions are common in IB phenomena, so are they common in the organizational sciences more generally. Moreover, they are not specific to any particular level of analysis. Cortina et al. (2015) gave examples at the country and organizational levels, but RV interactions can also be found at the between-unit level, the between-person level, and even the within-person level.

One of the great advantages of RV interactions is that they are relatively simple to conceptualize and defend. One need only explain why it is that values for one variable are implausible at certain levels of a moderator, and why at other levels of the moderator, values of the first variable are free to vary, and to covary with other variables. The classic example, and one with which many are in fact familiar, is the situation strength by personality interaction proposed by Mischel (1973) and reviewed by Meyer et al. (2010) in which the personality-behavior relationship is weak in strong situations because of the lack of variance in behavior in such situations. In weak situations, behavior is free to vary, and this variance is explained by personality.

This logic can be extended to many phenomena in the organizational sciences. For example, Ilies, Scott, and Judge (2006) hypothesized that intra-individual increases in positive affect would be associated with increases in organizational citizenship behaviors (OCBs), but

that this relationship would be weaker for those high in trait agreeableness. These authors argued that positive affect is necessary for those low in agreeableness to engage in OCBs. Those high in agreeableness engage in OCBs regardless of affect. Although Ilies et al. (2006) did not make RV arguments, their interactions could be framed in such terms. Y has less variance (i.e., it is uniformly high) when the moderator is high, in which case Y has less of a relationship with X. When the moderator is low, Y has more variance, and that variance is explained by X.

As we show, there are a great many other examples in management, and they exist at every level of analysis. Because an RV phenomenon is relatively easy to describe and justify, it is our contention that a broader appreciation of RV interactions would result in stronger justifications and tests of interaction hypotheses.

The purpose of the present paper is to explain and illustrate the nature and testing of RV interactions in organizational research. After a brief description of interactions more generally, we show the different forms that RV interactions can take and the sorts of arguments that should accompany them. We also differentiate RV interactions from range restriction phenomena, which are entirely different. We then detail the empirical evidentiary requirements of such interactions for a variety of designs and provide recommendations to guide future researchers in identifying, justifying, testing, and reporting RV interactions.

What are RV interactions?

Interaction effects describe a relationship in which the effect of predictor variable (X) on an outcome variable (Y) depends on the level of the moderating variable (Z), or more generally, the relationship between two variables is contingent upon some condition. At the least, a moderator variable affects the strength¹ of the X-Y relationship and thus helps researchers identify the boundary conditions under which a given relationship exists. More specifically,

moderator variables explain when (or for whom) an IV-DV relationship exists, or how strong it will be under different conditions (e.g., Andersson et al., 2014; Frazier, Tix, & Barron, 2004).

RV interactions are a specific type of interaction effect that stipulate a particular reason for the effect of the moderator on the X-Y relationship. More specifically, in an RV interaction, the *variance* of X or Y changes as a function of Z. In other words, the variance of X or Y is restricted (or enhanced) as a function of Z, and this restriction/enhancement of variance affects the magnitude (though probably not the sign) of the X-Y relationship. For example, Mischel (1973) postulated that the strength of a given situation constrains the behavior or cognitions in which an individual engages, potentially overriding the effect of personality differences:

Psychological "situations" and "treatments" are powerful to the degree that they lead all persons to construe the particular events the same way, induce *uniform* expectancies regarding the most appropriate response pattern, provide adequate incentives for the performance of that response pattern, and instill the skills necessary for its satisfactory construction and execution. [...] Individual differences can determine behavior in a given situation most strongly when the situation is ambiguously structured [...] so that subjects are uncertain about how to categorize it and have no clear expectations about the behaviors most likely to be appropriate (normative, reinforced) in that situation. [...] Conversely, when subjects expect that only one response will be reinforced... then individual differences will be minimal and situational effects prepotent. (p. 276)

To illustrate, consider the data in Tables 1a and 1b.

Insert Tables 1a and 1b about here

The data in Table 1a reflect a positive relationship between neuroticism and the expression of fear, and let us suppose that these data were collected in a “weak” situation. Let us

then ask what we would expect to happen if data for the same variables were collected in a situation in which there were strong norms against expressing fear (e.g., an Army platoon briefing, an NFL huddle). Presumably, the variance in a stable individual difference variable would not be influenced by the situation (we discuss the effects of selection in a later section), but the variance in expression of fear would be affected by norms. Specifically, the norms against fear expression would have the same sort of effect as would severity bias; suppressing all values, but suppressing large values more than small values. The data in Table 1b reflect this downward compression on the DV. The DV data in Table 1b were created by multiplying the DV values in Table 1a by .4. This transformation reduces large values more than it reduces small values, as would be the case for a phenomenon such as that described by Mischel. Nevertheless, the pattern of the X-Y relationship is the same at the two levels of the moderator in the sense that the rank order of Y relative to X remains unchanged, but higher values of the Y-variable are unlikely in the high strength condition because they have been compressed downward. As a result, Y has less variance in the strong condition than in the weak condition.

If we conduct these regressions separately, we get the results in Figures 1a and 1b. We see that the X-Y slope is 1.58 in the weak condition while it is .63 in the strong condition. It is important to note that the difference in these slopes is directly proportional to the differences in the standard deviations of Y at the two levels of the strength variable. The reduction in the standard deviation of Y by 60% results in a corresponding reduction in the covariance of 60%. Because the regression weight is equal to

$$b = \frac{COV_{xy}}{S_x^2} \quad (1)$$

the weight must also be reduced by 60%. Thus, the magnitude of the interaction is tied to the amount of restriction (or enhancement).

Insert Figures 1a and 1b about here

Now suppose that these two data sets were combined into a single data file along with a dichotomous moderator representing situation strength. A moderated regression (see Table 2) shows that the weight for the XZ product (-.945) is significantly different from zero. It is in fact negative because the X-Y slope is closer to zero in the strong condition than in the weak. These differences in weights are due to compression of the DV. This is the RV interaction effect.

Insert Table 2 about here

There are a number of different reasons how and why a moderator might restrict or enhance variance in one or more variables in a system. In the following section, we provide a variety of examples for the reasons that an RV moderator may affect the IV-DV relationship. The commonalities among these examples should illustrate the basis for the theoretical rationale that underlies an RV interaction.

Theoretical rationales for RV interactions

In order to explain how to make a case for an RV interaction, we begin by considering briefly how arguments are made for interaction effects more generally. In arguing for interaction hypotheses, one must make a case that a certain relationship is, in some way, variable, or as Cohen et al. (2003), Arnold (1982), Stone and Hollenbeck (1984), and others put it, conditional. Perhaps the simplest case is that in which Z is a categorical moderator of the X-Y relationship. If Z is truly categorical, then one can argue that the X-Y relationship is different for the different categories because of a stable difference between the categories.

If Z is continuous, then the case is more difficult to make because one cannot draw upon qualitative differences between groups (Andersson et al., 2014). In principle, one must argue that

the X-Y relationship changes as a function of the moderator. In a bilinear interaction, it must be argued that the X-Y relationship differs by a fixed amount for a given difference in Z. In a quadratic relationship, it must be argued that the X-Y relationship differs by a fixed amount for a given difference in X. For higher order interactions, it must be argued that lower order interactions differ by fixed amounts for a given difference in moderators, and so on.

Because interaction hypotheses are often coupled with main effect hypotheses, together they form a set that reflects a specific pattern. Defending specific patterns can be quite challenging. Cohen et al. (2003) provided labels and explanations for different combinations of main and interaction effects. Of particular interest for present purposes are the interactions described by Cohen et al. (2003) as *buffering* and *interference* interactions.

Buffering restricted variance interactions. RV interactions typically involve the weakening of an effect as the moderator changes in a particular direction, with values of Y converging on a certain value as the moderator moves in the “strong” direction. As we discuss later, this is not necessarily the case, but it is the most common form of RV interaction. One pattern that would result from an RV phenomenon would be that associated with a *buffering* interaction. In a buffering interaction, the moderator Z moderates the X-Y relationship because high levels of Z act as a buffer against the effects of X, reducing the importance of X for Y. The X and Z effects are of opposite sign, but the relationship between either and Y is weaker at higher levels of the other.

Cohen et al. (2003) explain that buffering interactions are common in research involving risk factors and protective factors. A certain physical (e.g., family history of a particular condition) or situational (e.g., low SES) characteristic is associated with undesirable outcomes (e.g., cancer, incarceration), but protective factors (e.g., diet, family support) act as a buffer. In

an RV phenomenon, one would argue that the relationship between the risk factor and the outcome is reduced because certain values of the outcome (high values in this case) are unlikely at higher levels of the protective factor. A buffering interaction is characterized by predictors that have weights of opposite sign, and increases in one predictor reduce the slope of the line representing the relationship between the other predictor and the outcome.

Mischel's situational strength arguments are buffering arguments (see Figures 1a and b). Personality drives behavior, but situational characteristics mitigate this influence such that situation strength makes certain behaviors unlikely. Consider our earlier example of neuroticism and fear expression. In informal, social situations, these two variables are positively related. In settings that impose a norm of (apparent) fearlessness, expression of fear is rare, and neither neuroticism nor anything else is likely to relate strongly to it. In terms of weights, we would expect a positive weight for neuroticism but a negative weight for strength. The product term would also have a negative weight because the slope representing the relationship between neuroticism and fear decreases as situational strength increases. Situational strength acts as a buffer against the effects of neuroticism on fear expression.

Another sort of buffering RV interaction comes from King and Ahmad (2010). These authors conducted a field experiment of the effects of Muslim identity of job applicants on discrimination directed at those applicants. They found that this relationship was positive (i.e., applicants wearing Muslim attire encountered more discrimination), but that it was weaker if the applicant exhibited warmth (which was stereotype disconfirming). Thus, warmth acted as a buffer against the effect of identity on discrimination, with the result being a pattern such as that in Figure 2, which is intended as a plot of the bar chart in Figure 2 of King and Ahmad (2010).

Insert Figure 2 about here

Our Figure 2 shows that the amount of discrimination experienced by Muslim applicants depends on whether or not the applicant exhibits warmth. On the other hand, we can see from King and Ahmad (2010) that discrimination was uniformly low for applicants who exhibited warmth. Warmth created a strong situation that kept discrimination low regardless of applicant attire. Only in the weak situation (no warmth information) was discrimination free to vary, in this case as a function of identity/attire.

As another example, Marcel, Cowen, and Ballinger (2017) examined the relationship between board use of an interim CEO and audience attitudes regarding board performance. They hypothesized that this relationship would be negative, but that it would be moderated by the suddenness of the departure of the previous CEO such that audiences (such as credit agencies, securities analysts, etc.) would react less negatively to the use of an interim CEO if the departure was sudden. This could be recast in RV terms. When the departure of the previous CEO is sudden, audiences are more sympathetic to the short term bind in which boards find themselves, so their attitudes toward board actions are (relatively) uniformly positive. When departure was not sudden, then attitudes have more variance, and this variance can be explained by reliance (or not) on an interim CEO.

Interference restricted variance interactions. A second category of interaction effect described by Cohen et al. (2003) is the *interference interaction*. In an interference interaction, the predictor and moderator have effects in the same direction, but the interaction works in the opposite direction. If the main effects are positive, then values of the outcome are restricted to high levels when either predictor is high. If the main effects are negative, then the opposite is true. In either case, the interference pattern is produced because values of the outcome variable converge upon a certain value at particular levels of one or the other predictor.

The Ilies et al. (2006) example given earlier is an example of an interference interaction. Agreeableness “interferes” with positive affect such that positive affect is not necessary for OCBs in highly agreeable people, for whom OCBs are high regardless of affect. As another example, Lyness and Heilman (2006) found that the positive relationship between performance ratings and probability of promotion were stronger for women than for men. Given that they also found that men received more promotions than women, this would be an example of an interference interaction in that the main effect of performance ratings is positive, men (coded as 1) were more likely to be promoted than women (coded as zero), and the performance rating-promotion slope was less positive for men, resulting in a negative weight for the product. The interaction is driven by the fact that variance in promotions is restricted when performance ratings are high (i.e., promotion is uniformly high), but much less restricted when performance ratings were low because men with low ratings were often promoted whereas women with low ratings were not.

At a higher level of analysis, Lu, Liu, Wright, and Filatotchev (2014) hypothesized that, in countries with strong market-supporting institutions, prior international experience is less important in determining Foreign Direct Investment (FDI) than it is in countries with weak market-supporting institutions. Strong market-supporting institutions help firms to identify and contact potential customers, navigate regulations, etc., making prior experience less important. Thus, FDI is uniformly high (i.e., low values are implausible) regardless of a firm’s prior international experience and is therefore only weakly related to it. In countries with weak market-supporting institutions, firms are left to fend for themselves. In such cases, FDI is low for firms with low levels of experience but high for firms with high levels of international experience. Thus, market supporting institutions interfere with the effects of experience. The

interaction between market-supporting institutions and experience can be justified simply by explaining why FDI is uniformly high where strong market supporting institutions are present and why FDI varies as a function of experience where such institutions are absent.

A very similar argument could be made in support of hypothesis 3a in Zheng, Singh, and Chung (2017). Where there is high capital market development, firms that wish to exit do so through sell-offs regardless of their political ties. In the absence of capital market development, firms may exit through sell-off or simple dissolution (i.e., there is variance in the manner of exit). Whether they sell-off or are merely dissolved depends on political ties.

In all of these examples, as a result of the interfering phenomenon, Y has less variance (i.e., it is uniformly high) when the moderator is high. When the moderator is low, Y has more variance, and this variance is explained by another variable. These examples also show that RV interactions can exist at any level of analysis.

Figure 3 contains data and results consistent with an interference RV interaction. Let us suppose that we hypothesize a model like that of Ilies et al. (2006). When agreeableness is low, OCB's are low or high depending on positive affect. High agreeableness, on the other hand, acts as a strong situation, compressing OCB values upward (i.e., low OCB values are implausible regardless of affect). The result is an affect by agreeableness interaction such that the weights for the predictors are both positive, but the affect-OCB relationship is weaker for those high in agreeableness.

Insert Figure 3 about here

In all of these cases, the argument for interaction can be simplified, or at least augmented, by explaining *why the DV varies little when the moderator is at a certain level*. Fear expression is unlikely in platoon meetings regardless of the personality of the subject. Discrimination is low

for applicants who exhibit warmth, regardless of attire. Promotions are high for high performers regardless of gender. OCBs are high for those high in agreeableness regardless of affect. FDI is high in countries with strong market supporting institutions regardless of firm experience. It is only when the “situation” is weakened (e.g., lack of norm for fearlessness, lack of warmth, low performance, low agreeableness, weak market supporting institutions) that substantial variance in the DV exists. This variance can then be explained by things like neuroticism, identity, gender, affect, and experience.

IV compression RV interaction. In all of the examples that we have offered, the RV interaction came about as a result of certain values of the DV being implausible at certain levels of the moderator. An RV interaction can, however, also result from compression on the IV. As we explain in a later section, this is entirely different from the restriction of range problem that is common in selection research (see also Aguinis et al., 2017). In any case, the result of IV compression is different from that of compression on the DV, and as we explain, these results are not widely appreciated.

Consider once again the examples in Tables 1a and 1b. Let us suppose that, for some reason, instead of wishing to predict fearful expressions from neuroticism, situation strength, and their product, we wished to predict neuroticism from fearful expressions, situation strength, and their product. Suppose further that the same situational factor was at work such that the values of the fear expression variable (now the predictor) were compressed downwards. If we then performed these regressions separately for each of the two levels of situation strength, we would have the plots in Figures 4a and 4b. As can be seen from these plots, the values of Y are uniformly low when X is low, just as is the case with Figures 1a and 1b. However, unlike Figures 1a and 1b, Figures 4a and 4b show that variability in Y doesn’t increase as a constant

function of X. Instead, variability increases until the (compressed) limits of X are reached, and then variability in Y drops back down.

Insert Figures 4a, 4b, 4c, and Table 3 about here

The results of the moderated regression, based on centered variables, are in Table 3. If we compare these results to those in Table 2, we see that the weight for the product is now positive, and the ΔR^2 is different. However, the results of the t-tests for the interaction are *identical* in the two examples. Thus, although compressing the values of the IV instead of the DV changed the *magnitude* of the interaction, the change in the standard error of the product was proportional, resulting in the same t and p values. If this were one's only interest, then it makes no difference whether the compression is in the IV or the DV.

On the other hand, the weight for the product indicates that the slope representing the relationship between Expressions of Fear and Neuroticism is *steeper* (i.e., more positive) in the strong situation. Recall from the regression of Fearful Expressions onto Neuroticism that the difference in weights was proportional to the difference in covariances produced by the compression. In the regression of Neuroticism onto Fearful Expressions, the covariances are affected in exactly the same way as they were in the earlier example. They must be because

$$COV_{xy} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{\sum(Y-\bar{Y})(X-\bar{X})}{N-1} = COV_{yx} \quad (2)$$

But because the compression in the present example occurs on the IV, the denominator of b is also affected. As per equation 1, the unstandardized regression weight is covariance divided by variance of X, and whereas the covariance in the strong situation is .4 of its value in the weak situation, the variance of X in the strong situation is .4² of its value in the weak situation. Thus, the denominator of b is reduced by a greater amount than is the numerator, and b is therefore smaller in the weak situation by a factor of .4²/.4=.4.

We had stated earlier that most RV interactions are either buffering (main effects of opposite sign) or interference (interaction effect of different sign from both main effects) interactions. This is because most RV interactions involve restriction (as opposed to enhancement) on the DV. If restriction occurs on the IV, however, the interaction may be of the synergistic sort, as in this example. All three terms in the regression have the same sign, meaning that the slope for one predictor gets more extreme as the other predictor increases.

We mentioned earlier that the effects of IV compression are poorly understood. It is interesting to note that the synergistic pattern described in the previous paragraph is the *opposite* pattern from that proposed by most authors of models in which IV restriction is implied. Consider the left-hand side of the model in Ozer (2011). OCBs directed at individuals and at organizations (OCB-I and OCB-O) are positively related to Team Member Exchange (TMX), but only if there is sufficient task autonomy: "...employees will be able to engage in OCBI [and OCBOs] to their coworkers and thus enjoy stronger TMX relationships with them only when they have sufficient freedom to perform such behaviors..." (p. 1329). Thus, the authors expect a positive main effect for OCB and a negative moderating effect of situation strength (i.e., lack of autonomy) owing to restricted variance in OCB when strength is high. This would be either a buffering or an interference interaction depending on the main effect of situation strength, but it wouldn't be a synergistic interaction.

If, in fact, OCBs are positively related to TMX but are uniformly low when situation strength is high, then we would actually expect the OCB-TMX slope to be *steeper* when situation strength is high, rather than shallower as the authors seem to propose. This type of confusion about RV interactions is very nearly universal as it is generally assumed in the literature that the

effects of RV would be the same regardless of where in the model the restriction occurs. As we show, however, this is not the case.

As another example, consider the literature on the Ideal Employee Factor (IEF: Schmit & Ryan, 1993). Schmit and Ryan (1993) showed that personality measures administered to non-applicants exhibited the expected five-factor structure, but measures administered to job applicants exhibited a sixth factor related to the desire to appear as the “ideal employee”. These authors suggested that employees who are motivated to do so (i.e., applicants) tend to inflate their responses, particularly to items whose relevance to the workplace is obvious. Schmitt et al. (under review) suggest something similar as a result of the ISF. If this were true, then we would expect upward compression of personality scores for applicants. If personality were used as a predictor of GPA (e.g., Schmit, Ryan, Stierwalt, & Powell, 1995), we would expect a *synergistic* interaction such that the positive relationship between, say, conscientiousness and GPA is stronger in the strong (i.e., applicant) condition because of the effects of compression on the covariance relative to its effects on the variance of Conscientiousness.

Of course, this means that *enhancement* on the IV also has counterintuitive effects. Let us once again consider the regression of neuroticism onto expressions of fear, but let us begin with the data in Table 1b. Suppose that a third variable *enhances* the variance of fear expression, and all values are increased by a factor of 2.5. Thus, large values are increased by a greater amount than are small values, and we have the data in Table 1a. This results in an increase in covariance from 5.78 to 14.44 (a factor of 2.5). s^2_x , however, increases by a factor of $2.5^2=6.25$, from 5.26 to 32.89. Thus, this enhancement on X increases the denominator in equation 1 by a greater amount than the numerator (2.5^2 rather than 2.5), and the b value is therefore *reduced* by a factor of $\frac{1}{2.5} = .4$, in this case from 1.098 to .439.

Restriction on both the IV and the DV. It should be clear from our previous discussion that restriction on both the IV and the DV by the same factor will not produce an interaction. In such a case, the covariance and the variance of X change by exactly the same amount, and b is unaffected. This can be seen by first considering the effect of compression on the numerator of equation 2. Compression would reduce deviation scores for both X and Y, with corresponding reductions in deviation score cross products. If one then considers the effect of compression on the denominator of equation 1, one sees that squared deviation scores for X would be reduced by the same amount. Thus, the numerator and denominator of equation 1 would be reduced by the same amount, and b would remain the same.

Consider the first stage of the mediation model presented by Ozer (2011). We quoted from p.1329 of that paper in order to make a point about compression of the IV, but the entire quote from Ozer (2011) is, "...employees will be able to engage in OCBI's [and OCBO's] to their coworkers and thus enjoy stronger TMX relationships with them only when they have sufficient freedom to perform such behaviors *and engage in social exchange with their coworkers* (p.1329, italics added)". Thus, the author suggests that situation strength compresses values of *both* OCB and TMX. If the compression effect is similar for these two variables, then in fact situation strength should not moderate their relationship.

If the restricting phenomenon affects the IV and DV by different amounts, b will differ by an amount proportional to the difference in factors by which the IV and DV variabilities are reduced. Our theories are generally not strong enough to suggest point (as opposed to directional) hypotheses, so it is unlikely that an author could base hypotheses on specific differences in compression. We simply wish to make clear that if compression of a certain amount is expected on *both* IV and DV, then no interaction should result.

Three-way RV interactions. RV phenomena also produce triple product interactions. Consider once again Figure 3, which contains an interference interaction intended to mimic the hypotheses of Ilies et al. (2006). OCB is uniformly high for those high in agreeableness but varies with positive affect for those low in agreeableness. But what if there were a unit or organization level variable that trumped inter- or intra-individual level phenomena? Suppose that the norms for citizenship behavior were very strong in one unit of an organization. In such a unit, OCB might be uniformly high regardless of positive affect *and* agreeableness, and we would see a plot like that in Figure 5.

 Insert Figure 5 about here

Taken together, these plots suggest that variance in OCB is low if *either* norms or agreeableness are high. Norms could be said to *interfere* with the affect by agreeableness interaction by compressing OCB values regardless of either. Only if both norms and agreeableness are low is OCB free to covary with intra-individual factors such as affect. We might label this a *second-order interference RV interaction*. (See Hypothesis 4a of Zheng et al., 2017 for another example).

Nonlinear and other RV forms. RV phenomena can also produce nonlinear effects. A *ceiling effect* is an RV phenomenon such that large values, usually of the outcome variable, are implausible. The true score of a given entity can only be observed if it is less than or equal to the “ceiling threshold” (Wang, Zhang, McArdle, & Salthouse, 2008). If a predictor is positively related to such a variable, then one will observe artifactual nonlinearity such that the X-Y relationship weakens when values of Y begin to hit its threshold, the reason being that high and very high values are compressed, thus restricting the variance of Y at higher levels of X. The result of such a phenomenon would be a scatterplot such as that in Figure 6.

 Insert Figure 6 about here

This form of variance restriction results in a positive weight for X but a negative weight for the polynomial term. This is a form of buffering effect in that the ceiling acts as a buffer of the X - Y relationship. Floor effects operate similarly. The true score can only be observed if it is greater than or equal to the threshold value, at which point any negative relationship with this variable would weaken².

Nonlinear RV phenomena can be more complicated still. The “glass ceiling” effect in the workplace is a form of RV *nonlinear* interaction. For men, better credentials or experience are associated with more advancement. For women, this may be true up to a point, but then a ceiling effect appears, rendering higher outcome variable values unlikely. Thus, for men, the relationship is linear and positive, whereas for women, the relationship is positive but concave downward because of restricted variance at high levels of X . Put another way, the ceiling buffers the effect of credentials for women, thus limiting the variance of advancement for qualified women, but not for qualified men. RV reasoning such as this could have been used to augment the arguments of Cotter, Hermsen, Ovadia, and Vanneman (2001) for glass ceiling effects for women. This reasoning also would have suggested a particular analysis strategy, beginning with a moderated regression (or random coefficient model in the case of Cotter et al., 2001) that included a $gender * experience^2$ term.

Conditional indirect effects containing RV interactions. Conditional indirect effect (CIE) models (also known as moderated mediation models or Models integrating Moderation and Mediation (MiMMs)) have become increasingly popular over the past 20 years. Such models are mediation models in which at least one path is moderated (Preacher, Rucker, and Hayes,

2007). These models, too, can result from RV phenomena. The multivariate nature of such models creates some unique RV issues.

Edwards and Lambert (2007) distinguished among several types of CIE or MiMM. All of the models discussed in Edwards and Lambert (2007) share the same $X \rightarrow M \rightarrow Y$ core but vary in the location of the moderation. Holland et al. (2017) found that the most common is the first stage ($X \rightarrow M$) moderation model. Consider as an example a portion of the model presented in Zhang and Bartol (2010). These authors argued that psychological empowerment mediated the effect of empowering leadership on intrinsic motivation. Furthermore, empowerment role identity (the degree to which one views oneself as wanting empowerment) moderates the relationship between empowering leadership and psychological empowerment. The authors argue that although empowering leadership and empowerment are positively related, there are those who have no interest in being empowered. For such people, empowering leadership is less likely to lead to empowerment. This argument could be recast in RV terms. For those with low empowerment role identity, empowerment is low (and therefore has little variance) regardless of leadership. For those high in empowerment role identity, empowerment varies, and this variation can be explained with empowering leadership. As a result of this first stage RV interaction, the indirect effect of empowering leadership on intrinsic motivation is stronger for those high in empowerment role identity. This argument requires only an explanation for why empowerment is uniformly low for some people, why leadership explains variance in empowerment for others, and why empowerment then leads to motivation.

As discussed earlier, RV effects differ depending on whether the restriction occurs in the IV or the DV. This is particularly important in CIE models because the mediator serves as the DV in one equation and as the IV in another. If the moderator restricts variance in X (the

exogenous variable), then this restriction has implications for all of the effects of X. In a partial mediation model with first stage moderation, RV in X must influence *both* the effect of X on M *and* the direct effect of X on Y. Thus, a mediation model such as that from Ozer (2011) *must* be what Edwards and Lambert (2007) call a direct effect and first stage moderation model rather than merely a first stage moderation model.

If instead the moderator restricts variance in M, then this has implications for *both* the first and second stages of the mediation. As we saw earlier, compression on the outcome variable in a relationship weakens its relationship with its predictor, but compression on the predictor variable strengthens its relationship with its outcome. Consider once again the Zhang and Bartol (2010) model. If variance in empowerment is low for those low in role identity, then there should be an empowering leadership by identity interaction as the authors hypothesize (i.e., X-M slope weakens as compression increases). In addition, however, there should be an identity by empowerment interaction on motivation such that the empowerment $\rightarrow$ motivation relationship is *stronger* for those low in role identity because of compression on what is now the IV. We might therefore expect a situation such as that depicted in Figure 7.

 Insert Figure 7 about here

When role identity is high (i.e., strength is low), the a and b coefficients in the indirect effect are .4 and .2, for an indirect effect of .08. When role identity is low, the coefficients are .2 and .4, resulting in the same indirect effect. Put another way, the compressing effect of role identity in one relationship is cancelled out by its compressing effect in the other, leaving the indirect effect unchanged across levels of the moderator.

As another example, consider the role of family logics in family-intensive models of governance in Miller, LeBreton-Miller, Amore, Minichilli, and Corbetta (2017). In this model,

family logics (i.e., community assumptions and values regarding centrality of family) lead to more family-intensive models of governance (i.e., family businesses continue to be run by family members: Hypothesis 1), and logics also moderate the relationship between family intensive models of governance and firm performance (Hypothesis 2) because family logics make it difficult to base hiring and promotion decisions on competence. Thus, X causes M and also moderates the relationship between M and Y because of its effect on Z. RV reasoning provides a simpler explanation.

In communities with strong family logics, family governance is more or less mandatory. Thus, top management is made up of family members regardless of competence of family members. “Familianness” of top management is high regardless of competence or anything else. In communities with weak family logics, “familianness” is not constrained to be high and is free to vary. It is also free to covary with competence of family members. The relationship between competence of family members and familianness is positive in communities with weak family logics, but it is near zero in communities with strong logics because familianness has so little variance.

In this model, familianness is a mediator between family competence and firm performance. Thus, familianness is the dependent variable in the first stage and the predictor in the second stage. Just as family logics reduce the competence-familianness relationship by compressing the variance of familianness, so must logics *strengthen* the positive relationship between familianness and firm performance by compressing variance on familianness. As a result, the indirect effect of family competence and firm performance is the same regardless of family logics. We found many examples of RV interactions stemming from mediator compression, but none in which this cancelling-out effect was appreciated.

Distinguishing RV interactions from range restriction. It is important to distinguish between RV interactions and issues relating to restriction of range. Such issues have been a common part of the personnel selection literature and are increasingly common in other areas of management (Aguinis et al., 2017). The term “restriction of range” is used to describe a situation in which there is direct or indirect selection on one of the variables in a system. The result is truncation, which biases parameter estimates and generally increases the probability of Type II errors. Consider the examples in Sackett and Yang (2000). Figures 1b and 1d from their paper represent the classic examples. In 1b, because of selection on X, the sample simply doesn’t contain any of the people who scored below a certain point on X. Figure 1d contains indirect restriction such that the probability of a person being excluded increases as a function of Z. Similar patterns can be found in strategy research because firms or joint ventures that perform very poorly are often omitted from common data sources (Bergh et al., 2014).

In RV phenomena, cases aren’t excluded. Rather, scores on a particular variable are compressed in a particular direction as a function of the moderator. The difference is best understood with the following description of truncation effects from Sackett and Yang (2000):

The larger reduction in sample covariance (the numerator in the formula for a correlation) relative to the reduction in a multiplicative term of two sample standard deviations s_x and s_y (the denominator in the formula for a correlation) leads to a *reduction* in the sample correlation from .50 to .33. (p.112, italics added for emphasis)

By contrast, the compression in an RV phenomenon creates *proportional* changes in the numerator and the denominator of the correlation formula, leaving r and β unchanged.

On the other hand, Sackett and Yang (2000) also note that selection/truncation on X has no effect on the unstandardized weight b . But as we showed earlier, the compression in an RV

phenomenon does affect b , and this is true regardless of whether the compression is on X or Y . This is because the denominator in the formula for b is s_x^2 , whereas the denominator of the formula for β is $s_x s_y$. As can be seen, the arguments needed to support RV interactions are very different from those needed to support a range restriction hypothesis because the latter involves *exclusion of cases* whereas the former involves *reduction in plausibility* of values of certain variables as a function of the moderator.

A related issue has to do with the distinction between differential prediction and differential validity. *Differential prediction* refers to unstandardized slope differences and is appropriately tested by moderated multiple regression or Latent Moderated Structural Equation Modeling (LMS, Klein & Moosbrugger, 2000). Evidence for differential prediction comes from a comparison of unstandardized regression coefficients at different levels of the moderator. *Differential validity* refers to correlational differences between the predictor and outcome (i.e., the percentage of variance in Y that is explained by X changes as a function of Z) and is usually tested by comparing correlation coefficients (or beta weights) across subgroups (Andersson et al., 2014; Carte & Russell, 2003; Arnold, 1982).

RV interactions represent a form of differential prediction. Standardized indices, like correlation coefficients and beta weights, are not affected by compression of values because they are based on variables with the same standard deviation. Put another way, RV affects covariance (the numerator of the correlation formula) and the product of standard deviations (the denominator of the correlation formula) proportionally, so standardized indices of covariation are unaffected. However, unstandardized indices, such as unstandardized regression coefficients and covariances, are influenced by compression of values. Thus, one can find an interaction effect due to compression of values even when there are no differences in correlations.

Consider once again the examples in Table 1. As we showed earlier, the compression of Fear Expression scores in 1b results in covariances and unstandardized weights that differ, and this produces the interaction in Table 2. The correlation, however, is the same in the two subsamples. Arnold (1982) demonstrates how standardization of variables can mask interaction effects by using as an example the relationship between a rectangle's length and area. Obviously, the area of a rectangle is the product of its length and width, i.e., they interact to determine area. Given two groups of rectangles of equal average length but different widths, if we regressed area onto length in each of the groups, we would find different unstandardized weights and, therefore, a length by width interaction. Yet the correlation between length and area is the same ($r=1$) in both groups, making it appear as though the relationship between length and area does not depend on width. Arnold (1982) goes on to state that:

[...] the fact that area is equal to the product of length and width causes σ_y to vary for groups of rectangles of differing widths, and these σ_y differences result in the necessary b differences which signal the fact that the form of the relationship between length and area is conditional upon width. (p. 148)

The primary implication of all of this is that although the typical tests for interaction such as moderated regression, LMS, or even multiple groups analysis with unstandardized values will detect an RV interaction, a comparison of correlations or standardized weights will not.

Issues in Testing and reporting RV interactions

The options for testing interaction effects generally depend on whether the moderator variable is categorical or continuous. When the moderator is categorical, the relationship between X and Y can be tested at the different levels (or within different groups) of Z in a multiple groups analysis. When the moderator is continuous, this option no longer exists without

losing information. Interaction effects are generally tested by forming products of predictors and examining the relationship between the partialled product and Y. In multiple groups analysis, an interaction effect is supported when the path linking X and Y is substantially different across the different Z groups. For continuous moderating variables, an interaction hypothesis is supported when the coefficient attached to the product term is substantial, showing that the product term contributes to the prediction of Y over and above the additive model. An increasingly common alternative is Latent Moderated Structural Equation Modeling (LMS, Klein & Moosbrugger, 2000) which estimates the relationship between the moderator and the X-Y slope without the use of products. Regardless of the analysis approach, the form or pattern of an interaction is demonstrated by reporting or plotting the X-Y relationship across different levels of Z (Cohen, Cohen, West, & Aiken, 2003, p.268).

The RV interaction is, above all else, an interaction. Tests of RV interactions begin with the same steps as do other sorts of interactions. As with most tests of interaction, one might begin with a moderated regression, looking for a substantial weight for the partialled product. If one were testing a cross level model, one might instead regress level-k slopes onto a level-k+1 moderator. If one were using Latent Moderated Structural Equation modeling (LMS), one might conduct the log-likelihood difference test between the linear and interaction models. There are, however, some additional requirements and some issues that are peculiar to the testing of RV interactions. In an RV interaction, the reason for the pattern is that certain values of one variable are implausible at certain combinations of values of another variable and the moderator. Thus, it must be demonstrated that variance in the restricted variable differs across levels of the other variable and/or moderator as hypothesized. To confirm the hypothesized RV pattern, one must

look at all of the coefficients (including the intercept) in the MR equation on centered variables, and one must also determine the actual pattern of heteroscedasticity as we explain below.

RV-specific follow ups. If one hypothesizes an RV interaction, then there are some specific follow analyses that are needed. Once the presence of an interaction has been determined, one must discover whether the interaction is due to compression as hypothesized³.

Let us return to Ilies et al. (2006) for an example. Agreeableness moderates the relationship between positive affect and OCB such that the relationship is weaker for those high in agreeableness. The RV explanation is that those who are high in agreeableness are uniformly high in OCB. As a result, positive affect is less predictive of OCB when agreeableness is high. Our analyses must of course show the appropriate affect by agreeableness interaction. In order to support the RV hypothesis, they must also show less OCB variance when agreeableness is high than when low.

As is usually the case with interactions, plots can be useful. As was mentioned earlier, Figure 3 contains a scatterplot that would be consistent with Ilies et al. (2006) with compression of Y when Z (or X) is high. Figures 8a and 8b also depict interactions that are consistent with the notion that the relationship between OCB and one exogenous variable is weaker when the other exogenous variable is high, but neither is consistent with this specific RV argument because the compression occurs in the wrong place. In 8a, the compression occurs in the middle of the agreeableness range. In 8b, it occurs at the beginning.

 Insert Figures 8a and 8b about here

In addition, it is useful to demonstrate that variance does in fact differ substantially across levels of the moderator (or the predictor). The most common tests for equality of variances are Bartlett's (1937) test and Levene's test (1960), which was developed to be less arduous and less

sensitive to departures from normality. Because Bartlett's test is less useful when raw data are available, we focus our attention on Levene and its common extension, Brown-Forsythe.

Levene's test is a one-way, fixed effects ANOVA on within-group, absolute deviations (Glass, 1966). Thus, it is an F-test against the null hypothesis,

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots \sigma_j^2$$

for j groups on $[(j-1), (n_1-1+n_2-1+\dots n_j-1)]$ degrees of freedom. Brown and Forsythe (1974) found Levene's test to be sensitive to departures from normality and developed a more robust procedure based on deviations from the median rather than the mean. Levene's test is available in most major statistical software packages. Unfortunately, what is labeled Brown-Forsythe in most packages (e.g., SPSS, R) is not the test for variance differences but is instead a robust test of mean differences. (Go to www.josecortina.com for detailed illustrations of these procedures using various software packages). Using the data from Table 1 as an example, we can determine whether the strength of the situation compresses variance on expressions of fear by running both Levene's and Brown-Forsythe tests. When we do this, both tests are significant (Levene's $F = 5.281, p = 0.03$; Brown-Forsythe $F = 4.81, p = 0.04$) thus allowing us to reject the null hypothesis that variance is equal across these two levels of the moderator. The difference in variances is in the anticipated direction, with less variance in the "strong" situation.

For continuous moderators, there are two options. The first is to categorize the moderator and conduct the same tests described in the previous paragraph. Although categorizing loses information, this can be a pragmatic solution. One must simply ensure that there are enough groups to detect any likely nonlinearities in the relationship between the moderator and variance (see Figure 8a for an example). In the (likely) absence of such nonlinearity, one might simply

select cases in the bottom and top quartiles on the moderator, and compare the two variances with Levene and/or Brown-Forsythe.

Alternatively, one might leave the moderator intact and conduct what amounts to a test for homoscedasticity. One of the assumptions of the OLS estimator is that, for any set of values of the predictors, the variance of errors is constant, which is to say, there is homoscedasticity. In testing an RV interaction, our interest is not in the variance of errors but in the variance of the compressed variable itself. If the compressed variable is the DV, then the standard test for constant error variances actually gives us what we need. In order to see this, one must first remember that there is no mathematical distinction between the “predictor” and the “moderator” in an interaction model – what is true of the moderator is true of the predictor. Next, one must realize that the standard test for homoscedasticity is equivalent to a test of the relationship between predictor and the variance of the DV (see p.73 of Berry, 1993 for a concise explanation of this fact). These two facts lead to the conclusion that the test for homoscedasticity provides the test for DV compression that we need.

The recommended test is that developed by Breusch and Pagan (1979). The first step in the Breusch-Pagan test is simply to regress the outcome onto the moderator with the goal of generating a column of residuals ($Y - \hat{Y}$ values). The residuals are squared, and the squared residuals are regressed back onto the moderator. A positive regression weight suggests that variance increases as the moderator increases, with compression of Y values at the low end of the moderator scale. Finally, the R^2 value from this last regression is multiplied by the sample size. This product is chi-squared distributed with degrees of freedom equal to the number of predictors. A significant chi-squared suggests that variance is not constant across values of the moderator (see the illustrations at www.josecortina.com).

If the compression is in the IV, then the equation can simply be turned around such that the IV is regressed onto the moderator, with the same procedure being otherwise followed.

DISCUSSION AND CONCLUSION

The purpose of this paper was to explore the nature of interactions that exist because certain values of one or more variables in a system are unlikely when another variable in the system takes on certain values. The result is compression, and this compression creates an interaction effect, the RV interaction.

We explain the nature of RV interaction arguments and make the case that such arguments are relatively easy to make, thus simplifying a justification process with which many struggle. In order to justify an RV interaction hypothesis, one need only explain a) that certain values of one or more variables in a system are implausible at certain levels of a moderator, thus restricting the variance of that variable, b) that the same variable has full variance at other levels of the moderator, and c) that these differences in variance create particular differences in relationships with other variables.

We then describe the various forms that such interactions can take. We show that, whereas restriction on the dependent variable tends to suppress its relationship with an independent variable, restriction on the independent variable has the opposite effect. We also discuss RV interactions in the context of mediation models, quadratic models, and higher order interaction models. We illustrate these forms with examples from the published literature and show how a given interaction form could be justified with a particular RV argument.

Finally, we discuss some of the empirical evidentiary requirements that are unique to RV interactions. Essentially, one must show that variance compression occurs where it is supposed to and nowhere else. Descriptions, examples, and requirements are summarized in Table 4.

Contributions to research

Our paper makes several important contributions. First, recent work has called for more appropriate theorizing of interaction effects, bemoaning that interaction effects are often not properly or inaccurately justified. More accurate theorizing comes from a better understanding of how interaction effects work. In the current paper, we demonstrate how researchers can use restricted variance logic and patterns to theorize about and test interaction effects. By offering concrete examples for appropriate theoretical justification, we hope to further theoretical development in a wide range of research that involves interaction hypotheses.

Furthermore, we explain how this class of interaction effects should be tested in order to determine whether findings are in fact consistent with underlying theory. We show how interpretations of prominent findings in the literature could be augmented with restricted variance arguments and with exploration of variance patterns in the data.

Identifying the precise dynamics of interaction effects may also have important practical implications. When we test specific RV models, we can better understand which conditions allow or exacerbate an effect or issue (such as stereotyping or discriminatory hiring and promotion) and which conditions may neutralize it, and why. Understanding the role of variance restriction could open up new avenues for research that illuminates how contextual factors can be employed to evoke desirable or suppress undesirable behaviors. At the same time, this knowledge can allow practitioners to create contexts that facilitate organizational interventions.

Finally, we hope that providing this introduction of RV interactions instigates a discussion of the importance of distinguishing different types of moderator models, their related theoretical rationales, and their testing. It seems that there is still much work to be done.

REFERENCES

- Arnold, H. J. (1982). Moderator variables: A clarification of conceptual, analytic, and psychometric issues. *Organizational behavior and human performance*, 29(2), 143-174.
- Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, 21(6), 1141-1158.
- Aguinis, H. (2002). Estimation of interaction effects in organizational studies. *Organizational Research Methods*, 5, 207-211.
- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *Journal of Applied Psychology*, 90, 94-107.
- Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, 20, 665-685.
- Andersson, U., Cuervo-Cazurra, A., & Nielsen, B. B. (2014). From the Editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies*, 45, 1063–1071.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 268-282.
- Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., & Joo, H. (2014). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal*, 37, 477-497.
- Berry, W.D. (1993). *Understanding Regression Assumptions*. Newbury Park, CA: Sage.
- Borack, J. I. (1994). Estimating predictive validity when range restriction due to selection and attrition is present. *Military Psychology*, 6(3), 193–204.
- Breusch, T. S.; Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica*. 47 (5): 1287–1294
- Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364-367.
- Carte, T. A., & Russell, C. J. (2003). In pursuit of moderation: Nine common errors and their solutions. *Mis Quarterly*, 479-501.

- Chowdhury, S. K., & Endres, M. L. (2010). The impact of client variability on nurses' occupational strain and injury: Cross-level moderation by safety climate. *Academy of management journal*, 53(1), 182-198.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Erlbaum.
- Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, 19(4), 915-922.
- Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, 6(4), 415-439.
- Cortina, J. M., Köhler, T., & Nielsen, B. B. (2015). Restriction of variance interaction effects and their importance for international business research. *Journal of International Business Studies*, 46(8), 879-885.
- Cotter, D. A., Hermsen, J. M., Ovadia, S., & Vanneman, R. (2001). The glass ceiling effect. *Social forces*, 80(2), 655-681.
- Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychological methods*, 12, 1-22.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of counseling psychology*, 51(1), 115.
- Glass, G. V. (1966). Testing homogeneity of variances. *American Educational Research Journal*, 3(3), 187-190.
- Holland, S.J., Shore, D.B., & Cortina, J.M. (2017). Review and recommendations for integrating mediation and moderation. *Organizational Research Methods*, 20, 686-720.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.)*. Thousand Oaks, CA: Sage.
- Ilies, R., Scott, B. A., & Judge, T. A. (2006). The interactive effects of personal traits and experienced states on intraindividual patterns of citizenship behavior. *Academy of Management Journal*, 49(3), 561-575.
- King, E. B., & Ahmad, A. S. (2010). An experimental field study of interpersonal discrimination toward Muslim job applicants. *Personnel Psychology*, 63(4), 881-906.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65(4), 457-474.

- Levene, H. (1960). Robust tests for equality of variances. *Contributions to probability and statistics, 1*, 278-292.
- Lu, J., Liu, X., Wright, M., & Filatotchev, I. (2014). International experience and FDI location choices of Chinese firms: The moderating effects of home country government support and host country institutions. *Journal of International Business Studies, 45*(4), 428-449.
- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology, 91*(4), 777.
- Marcel, J. J., Cowen, A. P., & Ballinger, G. A. (2017). Are disruptive CEO successions viewed as a governance lapse? Evidence from board turnover. *Journal of Management, 43*(5), 1313-1334.
- Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 97*(5), 951-966.
- Mendoza, J. L., & Mumford, M. (1987). Corrections for Attenuation and Range Restriction on the Predictor. *Journal of Educational Statistics, 12*(3), 282-293.
- Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36*(1), 121-140.
- Miller, D., LeBreton-Miller, I., Amore, M.D., Minichilli, A., & Corbetta, G. (2017). Institutional logics, family firm governance, and performance.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*(4), 252-283.
- Murphy, K.R., & Russell, C.J. (2017). Mend it or end it: Redirecting the search for interactions in the organizational sciences. *Organizational Research Methods, 20*, 549-573.
- Ozer, M. (2011). A moderated mediation model of the relationship between organizational citizenship behaviors and job performance. *Journal of Applied Psychology, 96*(6), 1328-1336.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: an expanded typology. *The Journal of Applied Psychology, 85*, 112-118.

- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78(6), 966.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607.
- Stone, E. F., & Hollenbeck, J. R. (1984). Some issues associated with the use of moderated regression. *Organizational Behavior and Human Performance*, 34, 195-213.
- Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate behavioral research*, 43, 476-496.
- Zhang, X., & Bartol, K. M. (2010). Linking empowering leadership and employee creativity: The influence of psychological empowerment, intrinsic motivation, and creative process engagement. *Academy of management journal*, 53, 107-128.
- Zheng, W., Singh, K.W., & Chung, C. (2017). Ties to unbind: Political ties and firm sell-offs during institutional transition. *Journal of Management*, 43, 2005-2036.

FOOTNOTES

1 – What we describe here is typically referred to as an ordinal interaction, that is, an interaction such that the magnitude of, say, the X-Y relationship changes as a function of Z, but the direction of the X-Y relationship is constant across plausible values of Z. Disordinal or crossover interactions are also possible, but RV interactions are, at least in principle, ordinal. Put another way, compression of values affects the magnitude but not the direction of slopes. Compression on Y pushes the numerator of the regression weight towards zero, but zero is the limit. Compression on X also pushes the numerator towards zero, but it pushes the denominator even harder towards zero, and the limit is an undefined value. Neither form of compression causes the numerator to change sign, and the denominator, because it is driven by squared values, can only be positive.

2 - As pointed out by an anonymous reviewer, although RV phenomena can produce ceiling/floor effects, such effects can result from many things including flaws in measures. In other words, an RV phenomenon such as that described by Cotter et al. (for women) will produce a scatterplot such as that in Figure 6, but a scatterplot such as this can also result from the use of measures that fail to capture all of the variation in a variable.

3 - Just as phenomena other than RV (e.g., flawed measures) can create a pattern such as that in Figure 6, so can other phenomena create other patterns that could be mistaken for RV. For example, a buffering RV pattern such as that is Figures 1a and 1b could also result from a positive relationship between X and the amount of measurement error in Y (e.g., if X is the degree to which a country is developed and Y is drawn from government records, X might be related to the quality of the measurement of Y (Berry, 1993).

Tables 1a and 1b

Data tables for weak vs. strong situations: relationship between neuroticism and the expression of fear

Table 1a. Data from an Unrestricted Variance Situation

Neuroticism	Expressions of Fear
1	6
2	2
3	8
4	4
5	10
6	8
7	12
8	20
9	12
10	18

Table 1b. Data from a Restricted Variance Situation

Neuroticism	Expressions of Fear
1	2.4
2	.8
3	3.2
4	1.6
5	4
6	3.2
7	4.8
8	8
9	4.8
10	7.2

Table 2

Moderated Regression Results with weights and R^2 at each step.

Predictors	b	Std. Error	t	p
Neuroticism	1.1	.225	4.89	.00
Sit. Strength	-6.0	1.29	-4.63	.00
$R^2=.728$				
Neuro * Strength	-.945	.40	-2.36	.031
$\Delta R^2=.07$				

Table 3

Results of the moderated regression of neuroticism on fearful expression by situation strength

Predictors	b	Std. Error	t	p
Fear. Express.	.53	.108	4.89	.000
Sit. Strength	3.18	1.11	2.87	.011
$R^2=.585$				
Fear * Strength	.659	.279	-2.36	.031
$\Delta R^2=.107$				

Table 4

Descriptions, examples, and evidentiary requirements for the different types of RV interactions.

RV interaction	Explanation	Example	Testing
DV Compression Buffering interactions	The moderator Z compresses the variance of Y in such a way that it acts as a buffer against the effects of X, reducing the importance of X for Y. The X and Z effects are of opposite sign.	Use of an interim CEO is negatively related to audience reactions, but sudden CEO departure acts as a buffer by making extremely negative reactions unlikely, thus reducing the negative relationship between use of interim CEO and reactions (Marcel et al., 2017)	A test for moderation is conducted, and the expected pattern is one in which the two main effects are in the opposite direction. Moreover, it must be demonstrated that DV variance changes as a function of the moderator in the expected direction. In this example, the Brown-Forsythe test should show that variance in audience reactions is significantly smaller for firms from which the CEO departed suddenly than for firms from which the CEO departed gradually.
DV Compression Interference interactions	The moderator Z compresses the variance of Y in such a way that it interferes with the effects of X, thus reducing the importance of X for Y. The X and Z effects are of the same sign, but the interaction is in the opposite direction. If the main effects are positive, then values of the outcome are restricted to high levels when the moderator is high. If the main effects are negative, then the opposite is true.	Agreeableness “interferes” with positive affect such that positive affect is not necessary for OCBs in highly agreeable people, for whom OCBs are high regardless of affect (Ilies et al., 2006).	A test for moderation is conducted, and the expected pattern is one in which the two main effects are in the same direction and the interaction is in the opposite direction. Moreover, the Breusch-Pagan test should show that variance in OCB decreases as a function of agreeableness.
IV Compression Interactions	The moderator Z compresses the variance of X. This <i>strengthens</i> the relationship between X and Y by reducing the numerator of b to a lesser extent than the denominator of b. In IV compression interactions, the variability in Y doesn’t change as a constant function of X. Instead, variability increases until the (compressed) limits of X are reached, and then variability in Y drops back down. Compression	For job applicants, self-reported personality scores are compressed upwards (e.g., higher conscientiousness), whereas for non-applicants, this is not the case (Schmit & Ryan, 1993). This compression would result in a stronger conscientiousness-GPA relationship for applicants than for non-applicants.	A test for moderation is conducted, and the expected pattern is one in which the X-Y slope is steeper (i.e., either <i>more</i> positive or <i>more</i> negative) when Z is at its stronger levels. Moreover, variance in self-reported conscientiousness should be smaller for applicants than for non-applicants.

	on the IV instead of the DV results in a different magnitude of interaction, however, the change in the standard error of the product is proportional, resulting in the same t and p values as would be found with compression on the DV. On the other hand, compression on the IV makes the X-Y slope steeper rather than shallower.		
Compression on both IV and DV	The moderator Z compresses the variance of both X and Y. If compression is equal, then there should be no XZ interaction. An interaction will exist to the degree that the amount of compression differs.	Ozer (2011) implied that a lack of autonomy would limit both OCB (X) and TMX (Y). If a lack of autonomy created the same amount of compression on both OCB and TMX, then there should be no OCB by autonomy interaction.	If equal compression is expected, then the interaction needn't be tested at all. One need only examine the effect of Z on the variances of X and Y. If unequal compression is expected, then an interaction would be expected. Regardless, the appropriate test of variance differences would be conducted on both variables.
3-way RV interactions	The moderator Z compresses variance on the DV in a twoway interaction. This compression on the DV will tend to weaken all relationships, both additive and multiplicative.	The interference interaction hypothesized by Ilies et al. might itself be interfered with if a unit-level variable compressed variance in OCB. OCB might be uniformly high in units with strong norms for citizenship. For units with weak norms, OCB might be uniformly high for those high in agreeableness, but vary as a function of positive affect for those low in agreeableness.	A test for three-way interaction is conducted, and the expected pattern is one in which, when the overriding variable (e.g., unit norms) is weak, the lower level IVs demonstrate their hypothesized interaction pattern. When the overriding variable is strong, most or all of the lower level relationships are washed out. Moreover, OCB variance, both within and between person, should be low in units with strong OCB norms. In units with weak norms, within-person variance in OCB should be low for agreeable people.
Non-linear RV effects and interactions	If an omitted moderator Z produces a ceiling or floor effect, the result will be a nonlinear relationship between X and Y due to changes in the variance of Y. If the ceiling/floor exists for	'Ceiling effect' refers to a phenomenon in which higher scores in a range are implausible. X and Y might be positively related, but if there is a ceiling for Y, then a diminishing	A test for a quadratic effect is tested, and the expected pattern would be concave downward or concave upward depending on whether there is a ceiling or a floor. If this curvilinearity is expected for some

	some people but not for others, then a nonlinear interaction is expected.	returns pattern will emerge such that higher X is associated with higher Y until the ceiling on Y is reached. Beyond that point, additional X is not associated with higher Y. The ceiling effect might exist for some people (e.g., women) but not for others. This would be the glass ceiling.	levels of a second predictor but not others (e.g., for women but not men), then a three-way ($X*X*Z$) interaction is tested. Moreover, it must be demonstrated that DV variance changes as a function of the DV itself (or of the IV as it amounts to the same thing). If there is a moderator of this relationship, then it must also be demonstrated that this variance function changes as a function of the moderator. Using the glass ceiling as an example, it must be demonstrated that promotion variance decreases as a person climbs the hierarchy and that promotion variance remains constant for men.
Conditional indirect effects with RV interactions	If moderator Z compresses variance on a variable in a CIE model, the effect depends on where in the model the compression occurs. Compression on X will strengthen both the direct and indirect effects of X on Y. Compression on Y will weaken both effects. Compression on M will weaken the X-M linkage but strengthen the M-Y linkage.	Zhang and Bartol (2010) hypothesized a model in which empowering leadership increases empowerment, which in turn increases motivation. If role identity compresses variance in empowerment, then increases in role identity should weaken the empowering leadership-empowerment relationship but strengthen the empowerment-motivation relationship.	A test for a conditional indirect effect is conducted such that the magnitude or direction of the relevant linkages is shown to vary as a function of the moderator, as is the indirect effect. Moreover, variance in empowerment should be low when role identity is low, thus weakening the first stage of the CIE but strengthening the second stage.

Figure 1a

Plot of the regression of Fear Expressions onto Neuroticism in a Weak Situation

Figure 1b

Plot of the regression of Fear Expressions onto Neuroticism in a Strong Situation

Figure 2

Replotting of the bar chart in Figure 2 of King and Ahmad (2010) in which warmth acted as a buffer against the effect of identity on discrimination

Figure 3

Contains data and results consistent with an interference RV interaction from Ilies et al (2006)

Figure 4a

Plot of the regression of Neuroticism onto Fear Expressions in a Weak Situation

Figure 4b

Plot of the regression of Neuroticism onto Fear Expressions in a Strong Situation

Figure 4c

Plot of the regression of Neuroticism onto Fear Expressions

Figure 5

Triple product RV interactions: Organizational Norms for Citizenship Behavior as a Strong Situation

Figure 6

Example of nonlinear RV interaction

Figure 7

Example of conditional indirect effects model containing RV interactions

Note: First path coefficients for each path are for when the compression on empowerment is low.

Numbers in parentheses represent path coefficients for each path for when compression on empowerment is high.

Figure 8a
Compression at moderate levels

Figure 8b
Compression at low levels

Appendix A: Categorical Moderators

To test for RV interaction when the moderator is categorical, one simply needs to determine if variance of the dependent variable differs across levels of the moderator. This can be done using equality of variance tests, such as Bartlett's, Levene's, or Brown-Forsythe. Below, we demonstrate Bartlett's test in Excel, R, and Mplus. Because the Bartlett's test cannot be conducted in SPSS without also using software such as Excel, we do not demonstrate Bartlett's in Excel. We then demonstrate Levene's and Brown-Forsythe in Excel and offer a strategy for conducting these tests with latent variable models. Because the test in SPSS and R that is labeled Brown-Forsythe is not in fact the Brown and Forsythe (1974, Journal of the American Statistical Association) test of equality of variances, we demonstrate these only in Excel. All of our Excel files can be found in the other Supplemental Materials.

Our illustrations are based on data examining how the relationship between extraversion and one's chattiness varies depending on the strength of the situation (i.e., at a funeral vs. at work).

How to Conduct Bartlett's Test

The Bartlett's test of equal variances generates a test statistic, B , which is approximately chi-squared distributed.

$$B = \frac{df_w \ln MS_w - \sum_j df_j \ln s_j^2}{1 + \frac{1}{3(k-1)} \left(\sum_j \frac{1}{df_j} - \frac{1}{df_w} \right)}$$

Here s_j^2 is the variance of group j , MS_w is the pooled variance (i.e, the sample size-weighted average variance) across all groups, df_j is $n_j - 1$, and df_w is $\sum(df_j)$. Bartlett's test is sensitive to departures from normality, but unlike other tests, it doesn't require raw data to calculate. For this reason, it is particularly useful for latent variable models with categorical moderators.

In Excel:

Step 1: Calculate mean and variance for each group.

Step 2: Calculate the degrees of freedom for each group variance (df_j) and for the pooled variance (df_w) as described above.

Step 3: Use values from Steps 1 and 2 to calculate the MS_w as $\frac{\sum(df_j * s_j^2)}{df_w}$

Step 4: Take the inverse of df_j and df_w

Step 5: Take the natural log of the variance for each group and of the MS_w

Step 6: Calculate df for the test itself as $k - 1$, where k equals the number of groups

Step 7: Calculate the numerator of the B statistic by subtracting the sum of the products of the natural log of the variance for each group and the df for each group from product of the df of the pooled variance and the natural log of the pooled variance MS_w .

Step 8: Calculate the denominator by subtracting the inverse of the df_w from the sum of the inverse of df_j and then multiplying this number by the inverse of three times df_B . Add one to this product.

Step 9: Calculate the *B* statistic by dividing the computed numerator by the computed denominator. Compare to a chi-squared critical value with *k*-1 degrees of freedom, where *k* equals the number of groups (cutoffs for $p < .05$ for 1, 2, and 3 *df* respectively are 3.84, 5.99, and 7.81. For moderators with more than 4 groups, please consult a chi-squared table).

Sample Output

	A	B	C	D	E	F	G	H	I	J	K
1		Weak	Strong								
2		6	2.4		BARTLETT'S TEST			Code Used			
3		2	0.8		df	1		=COUNT(B2:C2)-1			
4		8	3.2		B-num	6.68814402		=D13*D16-SUMPRODUCT(B13:C13,B16:C16)			
5		4	1.6		B-den	1.05555556		=1+1/(3*F3)*(SUM(B14:C14)-D14)			
6		10	4		B	6.33613644		=F4/F5			
7		8	3.2								
8		12	4.8		alpha	0.05					
9		20	8		p-value	0.01183023		=CHISQ.DIST.RT(F6,F3)			
10		12	4.8		B-crit	3.84145882		=CHISQ.INV.RT(F8,F3)			
11		18	7.2	Error (W)	sig	YES		=IF(F9<F8,"YES","NO")			
12	mean	10	4								
13	df	9	9	18							
14	1/df	0.11111111	0.11111111	0.05555556							
15	var	32.8888889	5.26222222	19.0755556				=SUMPRODUCT(B13:C13,B15:C15)/D13			
16	ln(var)	3.49313488	1.66055341	2.9484077				=LN(B15)			
17											

In R:

Within R, conducting the Bartlett's test is fairly straightforward using the following packages and syntax. Note that R syntax is case sensitive. Packages must be installed and active in order to run analyses.

```
library(stats)
bartlett.test(y~z,data=mydata)
```

```
Console ~/
> #Bartlett's
> bartlett.test(chatty~situation,data=chatty)

Bartlett test of homogeneity of variances

data: chatty by situation
Bartlett's K-squared = 6.3361, df = 1, p-value = 0.01183
```

In Mplus:

To conduct Bartlett's test for a latent variable model, one needs to obtain the latent variances for the compressed variable at each level of the moderator. This can be combined with a test for interaction by conducting a multiple groups analysis (MGA) with the moderator as the grouping

variable. After obtaining the latent variances from the MGA, use Excel to run the Bartlett's test as described above.

Our example comes from a data set that examines how the relationship between Perceived Organizational Support (POS), tenure, and Life Satisfaction (LS) differs between supervisors and non-supervisors. POS and LS are latent variables. Supervisor status is a categorical moderator.

Step 1: Run MGA for the different levels of the moderator

Step 2: Obtain the latent variances for the dependent variable

Step 3: Use the latent variances to conduct the Bartlett's test in Excel

```

TITLE:
    Bartlett's test in mplus example

DATA:
    FILE IS climate.dat;

VARIABLE:
    NAMES ARE tenure POS1-POS6 LS1-LS5 super;
    USEVARIABLES ARE tenure POS1-POS6 LS1-LS5 super;
    MISSING = ALL (-99);

    GROUPING = super (0 = nonsupervisor, 1 = supervisor);

ANALYSIS:
    TYPE = general;
    ESTIMATOR= ml;

MODEL:
    f1 BY LS1-LS5;
    f2 BY POS1-POS6;
    f1 ON f2
    tenure;

MODEL nonsupervisor:
    f1 BY LS1-LS5;
    f2 BY POS1-POS6;
    f1 ON f2
    tenure;

MODEL supervisor:
    f1 BY LS1-LS5;
    f2 BY POS1-POS6;
    f1 ON f2
    tenure;

OUTPUT:
    SAMPSTAT STANDARDIZED RESIDUAL TECH1 TECH4 PATTERNS;

```

By requesting TECH4 in the OUTPUT command, we obtain the latent variances for supervisors (0.737) and for non-supervisors (0.805). We then enter these values into the Excel formulas described above. As can be seen in cell G7, the Bartlett's test is not significant, indicating that the latent variances of Life Satisfaction do not differ between supervisors and non-supervisors. Thus, it seems that RV is not the reason for any interaction.

	A	B	C	D	E	F	G
1						BARTLETT'S TEST	
2						df	1
3						B-num	0.427759603
4						B-den	1.002393375
5		Supervisors	Non-supervisors	Error (W)		B	0.426738258
6	df	335	161	496		alpha	0.05
7	1/df	0.002985075	0.00621118	0.00201613		p-value	0.513593791
8	var	0.737	0.805	0.75907258		B-crit	3.841458821
9	ln(var)	-0.305167387	-0.216913002	-0.2756579		sig	NO
10							

How to Conduct Levene's and Brown-Forsythe Tests

The Brown-Forsythe test and the Levene's test for variance differences are both based on a one-way ANOVA conducted on *absolute values of deviation scores* using the formula below. Brown and Forsythe (1974) demonstrated that when the underlying distribution is approximately normal, it is appropriate to use group means to form the deviation scores, which is what Levene (1960) recommends. When the distribution is heavily skewed, then the median is more appropriate. Thus, in the Brown-Forsythe test, deviation scores are computed using medians rather than simple means, but both tests generate an F-statistic (often labeled as W) using the following:

$$F_{k-1, N-k} = \frac{N - k}{k - 1} \frac{\sum_{i=1}^k n_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - Z_{i.})^2}$$

Where

k = # of groups

n_i = # of cases in group i

N = total # of cases

Z_{ij} = the absolute value of the deviation score for case j in group i, $|y_{ij} - \tilde{y}_i|$

$\tilde{y}_i$ = either the mean or the median of y for group i depending on whether one is computing Levene's F or Brown-Forsythe's F

Z_{i.} is the mean of Z_{ij} for group i

Z_{..} is the mean of all Z_{ij}

In order to avoid confusion, we should mention that, although the df stipulated by Brown and Forsythe in the 1974 paper in which they developed their test for equality of variances (the JASA paper) are in fact, k-1 and N-k as we state above, some have suggested that the df in their *other* influential 1974 paper (in Technometrics, it's enough to make you cry) on testing mean differences be used instead. These df, usually attributed to Satterthwaite (1941, Psychometrika) and Welch (1951, Biometrika) are k-1 as before, and for the denominator, df are

$$\frac{1}{\sum_{i=1}^k \frac{m_k^2}{n_k - 1}}$$

Where

$$m_k = \frac{\left(\frac{n_k}{N}\right) * s_k^2}{\sum_{i=1}^k \left(\frac{n_k}{N}\right) * s_k^2}$$

Although there may be technical reasons to use these df, the difference in cut values for studies with adequate sample sizes is negligible. We suggest that one keep it simple and use k-1, N-k.

But enough about degrees of freedom. A simpler way to think about the Levene and Brown-Forsythe tests is as a oneway ANOVA on absolute deviation scores. This is demonstrated in Excel below.

RAW DATA			INPUT FOR ONEWAY ANOVA							
					Absolute deviation (Zij Weak)	Absolute deviation (Zij Strong)				
Weak	Strong	Y-Mean Weak	Y-Mean Strong							
6	2.4	-4	-1.6		4	1.6				
2	0.8	-8	-3.2		8	3.2				
8	3.2	-2	-0.8		2	0.8				
4	1.6	-6	-2.4		6	2.4				
10	4	0	0		0	0				
8	3.2	-2	-0.8		2	0.8				
12	4.8	2	0.8		2	0.8				
20	8	10	4		10	4				
12	4.8	2	0.8		2	0.8				
18	7.2	8	3.2		8	3.2				
Means	10	4								
Medians	9	3.6								
Var	32.88889	5.262222								

Anova: Single Factor						
SUMMARY						
	Groups	Count	Sum	Average	Variance	
	Absolute deviation (	10	44	4.4	11.37778	
	Absolute deviation (	10	17.6	1.76	1.820444	
ANOVA						
	Source of Variation	SS	df	MS	F	P-value
	Between Groups	34.848	1	34.848	5.280711	0.033766
	Within Groups	118.784	18	6.599111		4.413873
	Total	153.632	19			

The Levene's value of 5.28 shows that the variances are significantly different

The input for the oneway ANOVA is the two columns of *absolute deviation scores*. In this case, because we are conducting a Levene's test, these are absolute deviations from the mean. The ANOVA is a Single Factor ANOVA conducted with the Analysis Toolpak add-in (If you don't see Data Analysis under the Data tab, hold down (alt T I) and check off Analysis Toolpak). Of course, this ANOVA can be conducted with any statistical software, a fact that may be relevant for some Mac users.

As can be seen, the F value on 1 and 18 df, is significant, showing that the variances are significantly different at $p < .05$.

If one had reason to believe that there was substantial skew in the dependent variable in this analysis, then the Brown-Forsythe test is recommended instead. This test is identical to Levene's test except that deviations from the *medians* are used instead of deviations from the mean.

RAW DATA			INPUT FOR ONEWAY ANOVA					
			Y-Median		Y-Median		Absolute deviation (Zij)	Absolute deviation (Zij)
			Weak	Strong	Weak	Strong	Weak	Strong
	6	2.4	-3	-1.2			3	1.2
	2	0.8	-7	-2.8			7	2.8
	8	3.2	-1	-0.4			1	0.4
	4	1.6	-5	-2			5	2
	10	4	1	0.4			1	0.4
	8	3.2	-1	-0.4			1	0.4
	12	4.8	3	1.2			3	1.2
	20	8	11	4.4			11	4.4
	12	4.8	3	1.2			3	1.2
	18	7.2	9	3.6			9	3.6
Means	10	4						
Medians	9	3.6						

Anova: Single Factor

Groups	Count	Sum	Average	Variance
Absolute deviation (	10	44	4.4	12.48889
Absolute deviation (	10	17.6	1.76	1.998222

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	34.848	1	34.848	4.810897	0.041655	4.413873
Within Groups	130.384	18	7.243556			
Total	165.232	19				

The Brown-Forsythe value is, as usual, slightly smaller but still shows significantly different variances

The Brown-Forsythe F is also significant, suggesting that the variances differ.

WARNING: If one requests Levene's and Brown-Forsythe tests in SPSS or in R, one gets the correct Levene's value, but the value labeled Brown-Forsythe is *not* the test for equality of variances. Instead, it is a robust test for the difference between means from their 1974 Technometrics article. If one wished to generate Brown-Forsythe variance difference values in SPSS or R, one would have to generate the appropriate absolute deviation scores and then use these as input for a Oneway ANOVA. At that point, one may as well use Excel. Nevertheless, the SPSS syntax and R code for generating Levene's values are as follows.

In SPSS: With y as the compressed variable and z as the categorical moderator, use the following syntax:

ONEWAY y BY z
/STATISTICS HOMOGENEITY
/MISSING ANALYSIS.

In R:

```
library(car)
with(data, leveneTest(y, z, center=mean))
```

If one is not able to add the Analysis Tookpak, then the values for Brown-Forsythe F test can be computed directly as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2		Weak	Strong	Y-Median Weak		Absolute deviation (Zj Weak)	Zj weak - Mean Zj weak	squared	Y-Median Strong		Absolute deviation (Zj Strong)	Zj Strong - Mean Zj Strong	squared						
3		6	2.4	-3		3	-1.4	1.96	-1.2		1.2	-0.56	0.3136						
4		2	0.8	-7		7	2.6	6.76	-2.8		2.8	1.04	1.0816						
5		8	3.2	-1		1	-3.4	11.56	-0.4		0.4	-1.36	1.8496						
6		4	1.6	-5		5	0.6	0.36	-2		2	0.24	0.0576						
7		10	4	1		1	-3.4	11.56	0.4		0.4	-1.36	1.8496						
8		8	3.2	-1		1	-3.4	11.56	-0.4		0.4	-1.36	1.8496						
9		12	4.8	3		3	-1.4	1.96	1.2		1.2	-0.56	0.3136						
10		20	8	11		11	6.6	43.56	4.4		4.4	2.64	6.9696						
11		12	4.8	3		3	-1.4	1.96	1.2		1.2	-0.56	0.3136						
12		18	7.2	9		9	4.6	21.16	3.6		3.6	1.84	3.3856						
13																			
14	Means	10	4	Mean Zj	AVERAGE(F3:F12)	4.4	SUM(G3:G12)	112.4	Mean Zj	AVERAGE(K3:K12)	1.76	SUM(K3:K12)	17.984	Z, F12, K3:K12	AVERAGE(F3: F12, K3:K12)	3.08			
15	Medians	9	3.6	(Mean Zj-Z,)^2	(F14-\$N\$14)^2	1.7424			(Mean Zj-Z,)^2	(K14-\$P\$14)^2	1.7424							p-1	1
16				n*(Mean Zj-Z,)^2	F15*10	17.424			n*(Mean Zj-Z,)^2	K15*10	17.424						E16+I16	34.848	(N-1)/(p-1)
17	Var	32.8889	5.26222														G14+K14	130.384	18
18				Values in Yellow are for numerator															
19				Values in blue are for denominator															
20																			
21		16.4444	2.63111	19.07555556															
22		0.86207	0.13793																
23		0.74316	0.01902																
24		0.08257	0.00211	11.80811232		Welch's df													

Levene's F can be generated instead simply by replacing deviations from the median with deviations from the mean in columns F and K.

What if it is the predictor whose values are compressed? The analysis is the same. The inputted values are always those associated with the compressed variable regardless of where it is in a causal chain. By extension, if both the predictor and the outcome are hypothesized to be compressed, then each can be tested in turn.

What if the compressed variable is a latent variable? We recommend that users create scale scores from the indicators of the latent variable and follow the same steps in SPSS, Excel, or R using the scale scores. Again, one will need to convert the scale scores to absolute deviations from the mean or the median, and then conduct a one-way ANOVA.

Appendix B: Continuous Moderators

One possible way of handling a continuous moderator is to categorize it in some way (e.g., median split, top and bottom quartile) and then conduct one of the tests described above. A more elegant approach that maintains all of the information in the continuous moderator is to use the Breusch-Pagan test for constant error variance. This can be conducted directly in R. In SPSS and Excel, one generates residuals and then uses these to conduct the test. Mplus must be tricked into conducting the test.

How to Run the Breusch-Pagan Test in Excel and SPSS

Our illustrations are from data used to generate Figure 6 of the manuscript.

Step 1: Regress the dependent variable onto the moderator and save the residuals

Step 2: Square the residuals

Step 3: Regress the squared residuals onto the moderator

Step 4: Multiply the R^2 value obtained in Step 3 by the sample size

Step 5: This value is approximately chi-squared distributed with k degrees of freedom, where k is equal to the number of predictors.

Excel output

	A	B	C	D	E	F	G	H	I	J	K	L
1	Mod	DV	Residual	Squared Residual								
2	0.9	2.7	0.48899279	0.23911395								
3	2.4	4.7	1.56314027	2.44340751								
4	3.6	4.1	0.22245826	0.04948768		Regression Analysis				****Based on Figure 6		
5	4.3	3.3	-1.0096062	1.01930478		OVERALL FIT						
6	5.8	4.5	-0.7354588	0.5408996		Multiple R	0.41116872					
7	5.9	6.2	0.90281773	0.81507986		R Square	0.16905972					
8	7.7	6.9	0.49179471	0.24186204		Adjusted R Square	0.16177077					
9	8.4	6.9	0.05973021	0.0035677		Standard Error	2.47580943					
10	9.6	7	-0.5809518	0.337505		Observations	116					
11	10	7	-0.8278458	0.68532869								
12	1.7	3.7	0.99520478	0.99043255		ANOVA				Alpha	0.05	
13	3.3	5.7	2.00762876	4.03057325			df	SS	MS	F	p-value	sig
14	3.7	2.8	-1.1392652	1.29792529		Regression	1	142.170536	142.170536	23.1939746	4.5415E-06	yes
15	4.3	2.2	-2.1096062	4.45043853		Residual	114	698.778084	6.12963232			
16	5.4	6.3	1.31143524	1.71986239		Total	115	840.948621				
17	6.3	6.3	0.75592373	0.57142068								
18	7.8	7	0.53007121	0.28097549			coeff	std err	t stat	p-value	lower	upper
19	8.2	7	0.28317721	0.08018933		Intercept	6.89100504	0.30123077	22.8761656	2.0406E-44	6.2942692	7.48774088
20	9.3	7	-0.3957813	0.15664284		Squared Residual	-1.1298143	0.2345954	-4.8160123	4.5415E-06	-1.594546	-0.6650826
21	10	6.9	-0.9278458	0.86089785								
22	0.7	1.5	-0.5875602	0.345227								
23	2.5	3.3	0.10141677	0.01028536		Breusch-Pagan Test						
24	3.1	2.7	-0.8689242	0.75502933								
25	4.1	4.7	0.51384075	0.26403232		Chi square value	19.6109273			=R square*sample size		
26	5.6	5.7	0.58798824	0.34573017		df	1					
27	6.5	6.1	0.43247673	0.18703612		sig	YES					
28	7.4	6.8	0.57696522	0.33288886								

SPSS syntax and output

```
***Regress the dependent variable onto the moderator variable***
```

REGRESSION

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT DV
/METHOD=ENTER Mod
/SAVE RESID.
```

```
**Compute the squared residuals**
```

```
COMPUTE sq_res=RES_1*RES_1.
EXECUTE.
```

```
***Regress the moderator variable onto the squared residual***
```

REGRESSION

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Mod
/METHOD=ENTER sq_res.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.411 _a	.169	.162	2.4758

a. Predictors: (Constant), sq_res

ANOVA_b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	142.171	1	142.171	23.194	.000 _a
	Residual	698.778	114	6.130		
	Total	840.949	115			

a. Predictors: (Constant), sq_res

b. Dependent Variable: Mod

Coefficients_a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.891	.301		22.876	.000
	sq_res	-1.130	.235	-.411	-4.816	.000

a. Dependent Variable: Mod

Multiplying R^2 by the sample size (i.e., $0.169 \times 117 = 19.61$) gives the chi-square value.

How to run the Breusch-Pagan test in R

```
library(lmtest)
bptest(lm(y~z, data=mydata))
```

```
> ### Breuch-Pagan test for constant variance
> ## Using the bptest command in the lmtest package:
> bptest(lm(DV~Mod,data=Figure6data))
```

studentized Breusch-Pagan test

```
data: lm(DV ~ Mod, data = Figure6data)
BP = 19.611, df = 1, p-value = 9.492e-06
```

How to run the Breusch-Pagan test in Mplus

This example uses the same data as the latent variable example for Bartlett's test. In this example, age is the continuous moderator. POS is a latent endogenous variable with six indicators.

The Breusch-Pagan test is conducted using the syntax below:

```
TITLE:
Conducting the Breusch-Pagan test in Mplus;

DATA:
FILE IS rv example data.dat;

VARIABLE:
NAMES ARE id POS1 POS2 POS3 POS4 POS5 POS6 age;
USEVARIABLES ARE POS1-POS6 age;
MISSING = ALL(-99);

Constraint = age;

DEFINE:
standardize age;

ANALYSIS:

MODEL:
y BY POS1-POS6; ! Defines the latent endogenous variable.
y ON age; ! Regresses the latent dependent variable on the moderator.
y (resvar); ! Estimates the residual variance of y.
! Labels the residual variance of y as "resvar,"
! which allows it to be used in the MODEL CONSTRAINT command.

MODEL CONSTRAINT:
new(a b); ! Defines new intercept and slope parameters.
resvar = (a+b*age)**2; ! Constrains the residual variance of y to be equal
! to the regression of the squared residuals on the moderator.

OUTPUT:
SAMPSTAT;
```

To determine if there is support for non-constant factor variance, users examine the value of the “b” parameter created using the MODEL CONSTRAINT command. If this value is significant, then this indicates that the Breusch-Pagan test is significant (i.e., there is non-constant factor variance). In this example, the value for “b” is not significant, which means that the variance for the latent variable is not significantly different across levels of the moderator.

MODEL RESULTS					
		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Y	BY				
	POS1	1.000	0.000	999.000	999.000
	POS2	1.241	0.073	17.075	0.000
	POS3	1.079	0.067	16.115	0.000
	POS4	1.239	0.077	16.162	0.000
	POS5	0.510	0.091	5.626	0.000
	POS6	1.034	0.072	14.444	0.000
Y	ON				
	AGE	0.034	0.034	1.023	0.307
Intercepts					
	POS1	3.760	0.039	95.668	0.000
	POS2	3.622	0.044	81.807	0.000
	POS3	3.755	0.042	89.253	0.000
	POS4	3.609	0.047	77.025	0.000
	POS5	3.131	0.055	56.657	0.000
	POS6	3.608	0.044	82.111	0.000
Residual Variances					
	POS1	0.344	0.027	12.796	0.000
	POS2	0.323	0.029	10.989	0.000
	POS3	0.385	0.030	12.659	0.000
	POS4	0.438	0.036	12.278	0.000
	POS5	1.403	0.090	15.551	0.000
	POS6	0.505	0.037	13.663	0.000
	Y	999.000	0.000	999.000	999.000
New/Additional Parameters					
	a	0.617	0.026	23.218	0.000
	B	0.046	0.031	1.486	0.137