Organizational Research Methods
13(4) 668-689

DOI: 10.1177/1094428110380467
http://orm.sagepub.com
SAGE

The Presence of Something or the Absence of Nothing: Increasing Theoretical Precision in Management Research

Jeffrey R. Edwards[1] and James W. Berry[1]

Abstract

In management research, theory testing confronts a paradox described by Meehl in which designing studies with greater methodological rigor puts theories at less risk of falsification. This paradox exists because most management theories make predictions that are merely directional, such as stating that two variables will be positively or negatively related. As methodological rigor increases, the probability that an estimated effect will differ from zero likewise increases, and the likelihood of finding support for a directional prediction boils down to a coin toss. This paradox can be resolved by developing theories with greater precision, such that their propositions predict something more meaningful than deviations from zero. This article evaluates the precision of theories in management research, offers guidelines for making theories more precise, and discusses ways to overcome barriers to the pursuit of theoretical precision.

Keywords

philosophy of science, quantitative research, theory development

Theory testing in management research faces a paradox that confronts many social sciences. This paradox was articulated by Meehl (1967, 1978), who noted that in the hard sciences, such as physics and chemistry, improvements in research design lead to stronger tests of theories, subjecting them to increased risk of falsification. This phenomenon occurs because theories in the hard sciences can produce hypotheses that translate into predictions expressed as point values. As research designs become stronger (e.g., sample sizes are larger, measures have less error), estimates of point values have tighter confidence intervals, which increases the likelihood that hypotheses will be rejected. Hypotheses that survive these increasingly stringent conditions provide stronger corroboration of theories.

In the soft sciences, such as management, stronger research designs yield weaker tests of theories. This paradox arises because theories in the soft sciences usually express predictions not as point

[1] The Kenan-Flagler Business School, University of North Carolina at Chapel Hill, USA

Corresponding Author:

Jeffery R. Edwards, The Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Email: jredwards@unc.edu

values, but as directional statements, such as a positive or negative relationship between two variables. These predictions are tested by constructing a confidence interval not around some value predicted by the theory but instead around a null value indicating no effect. As research designs become stronger, confidence intervals around the null value become smaller, and the likelihood of concluding that the estimated effect falls on the predicted side of the null value approaches .50. To illustrate, suppose a theory predicts a positive relationship between two variables, and this prediction is tested with a sample of 380 cases and a p value of .05. Using a conventional two-tailed test, any correlation greater than .10 would be taken as support for the hypothesis. If a one-tailed test were used, taking into account the direction of the hypotheses, only 270 cases would be required to declare a correlation greater than .10 significant at $p < .05$. Hypotheses that accommodate such a broad range of values confront low hurdles of acceptance, and as a result, theories that such hypotheses are intended to test bear little risk of falsification.

The foregoing circumstances are widespread in management research. In study after study, hypotheses are framed as directional predictions, and support for hypotheses and their associated theories is inferred if some estimated parameter falls on the predicted size of zero. The questions being asked by these hypotheses are such that, if our studies were designed as well as could be imagined, an affirmative answer would boil down to a coin toss. Hence, it is little wonder that methodologically sound studies routinely claim support for the theories they test, and lack of support is often blamed on methodological flaws that, if corrected, would protect the theory from falsification rather than putting it at risk. This state of affairs creates a context in which theories are rarely laid to rest because their predictions have been rejected. Rather, we seem to repeat the pattern described by Meehl (1978, p. 807) in which "There is a period of enthusiasm about a new theory, a period of attempted application to several facet domains, a period of disillusionment as the negative data come in, a growing bafflement about inconsistent and unreplicable empirical results, multiple resort to *ad hoc* excuses, and then finally people just sort of lose interest in the thing and pursue other endeavors."

In methodological sources, the solution to this problem is often expressed as the recurring plea to abandon null hypothesis significance tests (Bakan, 1966; Carver, 1978; Cohen, 1994; Krueger, 2001; Nickerson, 2000; Rozeboom, 1960). Researchers who take this plea to heart might place confidence intervals around estimated parameters, documenting the degree of uncertainty surrounding an estimate. However, when drawing statistical inferences, researchers usually report whether the confidence interval excludes zero, which is tantamount to conducting a null hypothesis significance test. Reframing statistical tests in this manner does not resolve the paradox described by Meehl (1967, 1978). Rather, the root cause of the paradox, and hence its resolution, concerns the predictions generated by theories. If our theories yield nothing more than directional predictions, then a wide range of parameter estimates will be taken as support, and we will have little basis to revise or abandon the theories we test. However, if we reform the predictions we derive from theories, such that they predict the presence of something rather than absence of nothing, then stronger research designs will put our theories at greater risk, and tests of our theories can become more meaningful.

The purpose of this article is to promote theoretical precision in management research. To this end, we review theories sampled from the management literature and evaluate the extent to which the propositions set forth by these theories are precise. We then discuss ways in which theoretical propositions can be made more precise, broadening the array of possibilities beyond the requirement of point predictions described by Meehl (1967, 1978). Following this, we discuss ways to overcome barriers to theoretical precision in management research, highlighting behavioral and institutional forces to be addressed if we are to make progress toward the goal of theoretical precision.

The Precision of Theories in Management Research

We begin by assessing the precision of theories in management research. To conduct this task, we reviewed articles published during the past 25 years (1985–2009, inclusive) in the *Academy of*

Management Review, a primary outlet for theory development in management research. We identified the 20 most cited articles that were framed as attempts to develop theory, as opposed to taking stock of existing theory or discussing theory development itself. We evaluated the precision of each proposition set forth in each article by asking three questions. First, does the proposition specify the *magnitude* of a relationship? Propositions were considered imprecise if they merely stated the direction of a relationship (i.e., positive or negative) and were deemed increasingly precise if they predicted a range of values or, in the extreme, a point value. Second, does the proposition describe the *form* of a relationship? On this criterion, propositions were regarded as imprecise if they did not state the form of the predicted relationship, whereas propositions were considered more precise if the relationship was described as having a particular shape (e.g., linear, curvilinear, and logarithmic), with the greatest precision indicated by relationships for which specific features of the function were described (e.g., a curvilinear relationship with its peak at a particular point). Third, does the proposition identify *conditions* that influence the magnitude or form of the relationship? Imprecise propositions said nothing about factors that might impact the predicted relationships, whereas more precise propositions specified factors such as population, context, time, or other variables thought to affect the magnitude or form of the relationship. These criteria were initially applied by both authors to 5 of the 20 theories, and discrepancies in the evaluations were discussed and resolved. Once agreement was reached, the second author evaluated the remaining theories, and the final evaluations were reviewed with the first author to ensure consistency and accuracy.

In total, the 20 theories set forth 183 propositions. Regarding the magnitude of the relationships predicted by these propositions, 19 (10.4%) simply stated that a relationship would exist, 164 (89.6%) described the direction of the relationship, and none of the propositions predicted a point value or range of values. With respect to the form of the relationship, 177 (96.7%) of the propositions were silent on this issue. Of the 6 propositions that addressed form, 3 (1.6%) described the shape of the relationship, and an additional 3 (1.6%) articulated specific features of the relationship, referring to peaks or turning points in curvilinear functions. Finally, 44 (24.0%) of the propositions included conditions that influence the predicted relationship, most of which were cast as moderator variables, whereas 139 (76.0%) propositions stated predictions without reference to conditions.

The results of this assessment indicate that Meehl's (1967, 1978) observations about the soft sciences characterize theories in management research. For the most part, the theories in our sample developed propositions that predicted the direction of a relationship but said little about the form of the relationship or conditions that might influence the relationship. Rather, the majority of the propositions essentially stated that, if one variable increases, another variable will increase or decrease. It follows that, as the methodological rigor of studies designed to test these propositions increases, the likelihood of finding support for the propositions and their associated theories will likewise increase, putting the theories at progressively lower risk. We emphasize that this assessment is not meant to criticize the particular theories we examined. Rather, the conclusion we underscore is that even highly influential management theories, as represented by the most cited articles in what is arguably the premier theory development journal in the field of management, are susceptible to the paradox lamented by Meehl. This conclusion enjoins us to tackle theoretical precision in management research, determining how we can make theories more precise and, in conjunction, how we can overcome barriers that might hinder our progress toward theoretical precision.

Increasing Theoretical Precision

How can we increase the precision of management theories? The ideal ascribed by Meehl (1967, 1978) to the hard sciences might seem beyond reach, given that management theories rarely have the mathematical foundations necessary to generate point predictions. Nonetheless, precision can take less extreme forms that would nonetheless improve upon the directional predictions in management theories. In this section, we consider various approaches to increasing theoretical precision and discuss how these approaches can be applied in theory development.

Expand the Null Hypothesis

One way to increase theoretical precision is to treat the null hypothesis not as a single value but as a range of values that can be considered negligible from a theoretical standpoint (Binder, 1963; Cortina & Dunlap, 1997; Fowler, 1985; Hodges & Lehmann, 1954; Meehl, 1967; Murphy, 1990; Serlin, 1993; Serlin & Lapsley, 1985; Tryon, 2001). With this approach, theories would be required to predict not merely that a parameter differs from zero but that the parameter deviates from zero by some minimum threshold. The range of the null hypothesis can be framed as a "zone of indifference" that comprises "values which are essentially equivalent to the null hypothesis for our present theory or practice" (Binder, 1963, pp. 110-111). Values within range null hypotheses can be conceived as the "crud factor" that arises from the "obvious fact that everything is more or less correlated with everything in the social sciences" (Meehl, 1990a, p. 123) or the "ambient noise level" that represents the "average shared variance of 'unrelated' variables" (Lykken, 1968, p. 153). Theories that posit null hypotheses with wide ranges would be considered stronger than those that stipulate null hypotheses with narrow ranges, given that wider range null hypotheses reduce the set of empirical outcomes that would be considered consistent with the theory.

The boundaries of range null hypotheses can be specified using various approaches. One approach is to examine empirical relationships among variables that are independent from a conceptual standpoint. Using this approach, Lykken (1968) surmised that theoretically unrelated psychological variables share about 4% to 5% common variance, which translates into expected correlations ranging from .20 to .22 in absolute magnitude. Similarly, Meehl (1990a) reviewed various substantive domains in which measures designed to be unrelated exhibited correlations around .20 to .30. Results from meta-analyses could also be used to calibrate correlations among variables that should be conceptually unrelated as use this evidence to set boundaries around null hypotheses. Boundaries can also be drawn according to the substance and maturity of the theoretical domain under study. For instance, Tryon (2001, p. 379) discussed placing boundaries around null hypotheses that demarked values "considered inconsequential ... on the basis of substantive theoretical considerations." Likewise, Serlin (1993, p. 351) indicated that the width of a range null hypothesis "is determined by the state of the art of the theory and by the strengths of the auxiliary theories on which the prediction is based." Boundaries of range null hypotheses can also be based on practical considerations, representing deviations from zero that matter from an applied standpoint (Kirk, 1996; Spencer, 1995). In this vein, Murphy (1990, p. 404) suggest "if one was testing the effectiveness of a training program designed to reduce rating inflation on a 7-point evaluative rating scale, one might decide that the difference between the means for trained and untrained raters must be greater than .5 scale points to conclude that the training has had any real effect."

Boundaries for range null hypotheses can be difficult to justify. Indeed, such boundaries might even seem arbitrary, particularly when compared to the unequivocal value of zero usually stipulated by point null hypotheses. Certainly, setting boundaries for range null hypotheses requires researchers to make and defend important theoretical decisions. However, researchers are entrusted with numerous decisions at each stage of the research process. When commenting on judging effect sizes as trivial versus important, Kirk (1996, p. 755) observed:

> It is true that an element of subjectivity is introduced into the decision process when researchers make this kind of judgment. And the judgment inevitably involves a variety of considerations, including the researcher's value system, societal concerns, costs and benefits, and so on. However, I believe that researchers have an obligation to make this kind of

> judgment. No one is in a better position than the researcher who collected and analyzed the data to decide whether or not the results are trivial. It is a curious anomaly that researchers are trusted to make a variety of complex decisions in the design and execution of an experiment, but in the name of objectivity, they are not expected or even encouraged to decide whether data are practically significant.

As management researchers, we should accept the responsibility for judging how far from zero relationships must deviate to be theoretically meaningful. At a minimum, we should predict relationships that not only differ from zero but also rise above the ambient noise that characterizes the substantive domains of our research. As theories become more refined through conceptual development and empirical testing, we might expand the range of the null or calibrate deviations from zero in terms of theoretical importance. To be sure, setting ranges around null hypotheses requires careful and informed judgment, but the difficulty of this task should not compel us to throw up our hands and resort to point null hypotheses that are themselves hard to defend from a conceptual standpoint (Meehl, 1990a; Serlin & Lapsley, 1985).

Set Upper and Lower Limits

Theoretical precision can also be enhanced by setting upper and lower limits on parameters. Perhaps the most obvious limits are the statistical bounds of parameters that correspond to hypotheses. As noted by Meehl (1990a, p. 128):

> The Pearson *r* coefficient and its surrogates go from zero to one; analyses of variance and covariance are expressible in terms of proportion of variance accounted for; beta coefficients in a multiple-regression equation, the weights in a linear discriminant function, the factors in a factor analysis, the base rate and hit rates in taxometrics—all of which collectively comprise 90% of research in "soft" psychology—have mathematically defined ranges of possible values.

To illustrate, if the correlation between two variables reaches 1.00, then the variables are empirically redundant, and the constructs they represent should not be treated as distinct theoretical entities. This reasoning underlies the principle that correlations between latent variables should be significantly less than 1.00 to establish discriminant validity (Anderson & Gerbing, 1988; Bagozzi, Yi, & Phillips, 1991; Schmitt & Stults, 1986). Although it might seem obvious that variables treated as distinct should correlate less than 1.00, the management literature contains published studies in which correlations between measures of purportedly different constructs reach or exceed 1.00 when the correlations are corrected for attenuation using the reported reliabilities of the measures.

Requiring that parameters deviate from their minimum and maximum possible values raises a dilemma similar to that associated with point null hypotheses, in that we can assume a priori that any parameter will differ from any specific value. For instance, when seeking to establish discriminant validity, it is likely that the correlation between any two factors will differ from 1.00, even if that difference is miniscule. Moreover, parameters that approach but do not reach their minimum or maximum values might nonetheless be excessive from a substantive perspective, as when the correlation between purportedly distinct constructs hovers around .90 or .95. To resolve this dilemma, boundaries can be set inside the minimum and maximum limits of parameters, analogous to ranges around null hypotheses. Thresholds used to calibrate departures from minimum and maximum values would necessarily depend on the substantive domain of the constructs in question (Anderson & Gerbing, 1988), and researchers should exercise informed judgment when establishing these thresholds. Combining these thresholds with those that specify range null hypotheses would decrease the set of theoretically admissible values for a parameter, further increasing the precision of theories in the process.

Limits on parameters can be set using various criteria. Some criteria can be derived from the mathematical limits of parameters (Meehl, 1990a). For instance, the correlation between any two variables is limited by the correlations of each variable with a third variable. These limits are given by the following expression (McNemar, 1962, p. 167):

$$r_{12} = r_{13}r_{23} \pm \sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}, \tag{1}$$

where r_{12} is the correlation between the two variables under consideration and r_{13} and r_{23} are the correlations of these variables with a third variable. To illustrate, if r_{13} and r_{23} both equal .50, then r_{12} is bounded by $-.50$ and $+1.00$, as opposed to -1.00 and $+1.00$. As another example, if the effect of an independent variable on a dependent variable is fully transmitted through a mediator variable, then the relationship between the independent and dependent variables equals the product of the paths relating these variables to the mediator variable (Alwin & Hauser, 1975). If maximum values are specified for these paths, then the relationship between the independent and dependent variables is restricted to the product of the maximum values of the paths. Along similar lines, limits on parameters can be developed by considering constraints that result from the populations, contexts, and time frames that define the boundary conditions of a theory (Johns, 1991). Limits can also be based on conceptual reasoning, such as arguing that the relationship between constructs treated as distinct should not exceed some maximum value. Finally, limits can arise from pragmatic concerns, as when a potential predictor of employee performance is rejected if its multiple correlation with other predictors exceeds a certain level.

State Predictions as Comparisons

Theoretical precision can also be increased by stating predictions comparisons. This approach enhances precision not by predicting that an effect should differ from zero or fall within upper and lower limits, but instead by predicting that one effect will differ from another effect. Comparative predictions facilitate the pursuit of strong inference (Platt, 1964) in which rival hypotheses are pitted against one another and studies are designed such that evidence supporting one hypothesis necessarily refutes other hypotheses. Comparative predictions are involved when researchers test competing theories (Binder, 1963; Lakatos, 1978; Meehl, 1990a; Serlin & Lapsley, 1985) and evaluate alternative structural equation models (Anderson & Gerbing, 1988; Rodgers, 2010; Vandenberg & Grelle, 2009). Our focus here is on using comparative predictions to increase the precision of a single theory, such that a theory becomes more precise to the extent its predicted effects can be rank ordered in terms of magnitude.

Theories that state predictions as comparisons should include a good-enough belt that defines a range of differences between the predicted effects (Cohen, 1994; Tukey, 1991). This premise follows logically from the fact that predicting that one effect differs from another effect is equivalent to predicting that the difference between the two effects differs from zero. Just as we can assert that no effect is exactly zero, we can likewise claim that no two effects are exactly the same. As put by Tukey (1991, p. 100), "It is foolish to ask 'Are the effects of A and B different?' They are always different—for some decimal place."

Comparative predictions can be derived in several ways. For instance, previous research can be used to determine whether one effect is typically larger than another. This type of evidence can be drawn from individual studies or meta-analyses documenting the magnitudes of effects that pertain to the theory at hand. Comparative predictions can also result from logical arguments that explain why the effects occur. If one effect rests on a single explanation while another effect arises from several explanations that build on one another, then the latter effect is likely to be stronger than the former effect. Effects are also likely to be stronger when the variables involved are positioned closer to one another in terms of causal flow. For example, if one effect refers to a proximal cause of an outcome and another effect involves a distal cause of the outcome, such that the latter effect is implicitly mediated by several intervening constructs, then the former effect is likely to be stronger than the latter effect, given that mediated causal chains dampen the relationship between variables at either end of the chain. Comparative predictions can also be justified when conditions that bring about one effect are generally more prevalent than those that give rise to another effect.

Develop Nonnil Predictions

The foregoing approaches to increasing theoretical precision specify values from which parameters should differ. An alternative approach is to specify values that parameters should equal (Gigerenzer, 2004; Meehl, 1967, 1990b, 1997). With this approach, theorists would pursue the ideal that Meehl (1967) ascribed to the hard sciences, in which theories generate hypotheses expressed as nonzero values. This approach reframes how we develop and evaluate theories, such that theories are corroborated not by rejecting nil hypotheses that represent no effect (Cohen, 1994) but instead by failing to reject theoretically derived alternative hypotheses. Such hypotheses would require the researcher to specify prior to collecting data "*the magnitude of the population effect that will constitute theoretical support*" (Serlin, 1993, p. 351, emphasis in original). As put by Meehl (1990b, p. 231) in reference to psychological theories:

> Psychologists attempting to test a substantive theory in soft psychology should strive for a rationale by which an expected *amount* of effect could be predicted from the theory. Point values are ideal, but even in physics and astronomy they are surrounded by a tolerance based on an estimate of the experimental error. One hopes that, when enough persons become sufficiently skeptical about the weak corroboration provided by merely showing that the *X*s get higher scores than the *Y*s, that cheap and easy derivation might be replaced by one that says something about the range of non-null differences that would be consistent with the theory.... Because of the crud factor's ubiquity, merely saying that "there ought to be a difference between A and B" is a feeble test of anything, and we ought to work harder than we usually do to come up with some statement about points and ranges.

Nonnil hypotheses should be expressed as ranges rather than point values, given that hypotheses predicting any single point can be rejected a priori, regardless of whether that point is zero or some nonzero value (Meehl, 1990b; Serlin & Lapsley, 1985; Tukey, 1991). From a theoretical standpoint, nonnil hypotheses with narrower ranges can be considered stronger than nonnil hypotheses with wider ranges, given that narrower range nonnil hypotheses represent more precise predictions that are easier to refute, thereby putting the theory at greater risk of falsification.

Nonnil hypotheses can be derived in various ways. For instance, findings from previous studies can be used to establish hypothesized values for subsequent studies (Cudeck & Browne, 1983; Gigerenzer, 2004; MacCallum, Roznowski, Mar, & Reith, 1994). As argued by Mulaik, Raju, and Harshman (1997, p. 98), "A hypothesis one ought to test is that the effect is equal to the value estimated in the previous study, which one judged to be significantly different from a zero effect." This approach is consistent with principles of replication (Carver, 1993; Lykken, 1968) and cross-validation (Mosier, 1951; Snee, 1977) and characterizes what Anderson and Gerbing (1988, p. 412) call the "quintessential confirmatory analysis" in which the parameters of a model are constrained to previously estimated values (Cudeck & Browne, 1983; Gigerenzer, 2004; MacCallum et al., 1994). Nonnil hypotheses can also draw from meta-analyses, using population means to center nonnil values and population standard deviations to calibrate ranges around these values. Using meta-analyses for this purpose would help realize their potential for theory refinement (Schmidt, 1992). Nonnil hypotheses can also be based on procedures used in power analysis to determine expected effect sizes (Cohen, 1988; Lipsey, 1990; Murphy, Myors, & Wolach, 2009), which can be expressed as a ranges of values compatible with range nonnil hypotheses (Gillett, 1994). Nonnil hypotheses might also be derived through formal theorizing (Adner, Polos, Ryall, & Sorenson, 2009; Freese, 1980; Land, 1971), which can be pursued by developing analytical models, conducting simulations, and applying formal logic (Adner et al., 2009; Davis, Eisenhardt, & Bingham, 2007; Harrison, Lin, Carroll, & Carley, 2007; Hulin & Ilgen, 2000). Finally, practical concerns can be used to derive nonnil hypotheses, resulting in ranges of effect sizes that are considered important and useful (Fowler, 1985; Kirk, 1996; Spencer, 1995).

Specify Functional Forms

Theoretical precision can also be enhanced by specifying the functional form of a hypothesized relationship (Ferris et al., 2006; Guion, 1998; Meehl, 1978, 1990a, 1997). As indicated by our review, theoretical propositions are usually silent about functional form, implying that the proposed relationship is simply some monotonic function. Propositions can be made more precise by describing the form of the relationship, such as whether the relationship is linear, curvilinear, or piecewise linear (i.e., a function that has linear segments with different slopes). Describing the functional form of a relationship enhances theoretical precision by increasing the number conditions that must hold to support a hypothesis. For example, assume that multiple regression analysis was used to test various hypothesized functions. For a positive linear relationship, the coefficient on a first-order term should be greater than zero and, at the same time, coefficients on higher order terms should not differ from zero. For a quadratic function with its peak at a particular point, the coefficient on a squared term should be negative, the coefficients on additional higher order terms should not differ from zero, and the coefficient on the first-order term should take on a value that locates the peak of the function near the hypothesized point (Stimson, Carmines, & Zeller, 1978). More complex functions involve additional parameters that translate into further conditions that must be jointly satisfied to corroborate the hypothesized relationship. Predictions for parameters that describe functional forms should be accompanied by ranges that specify acceptable values, much like those that surround nonnil hypotheses.

Deriving propositions expressed as functional forms can be difficult, given that most management research is based on the implicit assumption that relationships among variables are linear. This assumption is manifested in meta-analyses that use correlations to represent effect sizes, given that a correlation is essentially a linear regression function between two standardized variables. The scarcity of evidence relevant to nonlinear relationships means that prior empirical research provides little basis for developing hypotheses that address functional form or departure from linearity. Nonetheless, certain streams of research have probed nonlinear relationships, and the conceptual logic underlying this research might be generalized to other areas of inquiry. These streams include research on diminishing marginal utility of economic and social goods (Diener, Ng, & Tov, 2009; Lane, 2000), risk-seeking and aversion (Tversky & Kahneman, 1992), negotiation (Northcraft, Preston, Neale, Kim, & Thomas-Hunt, 1998), stress (Ferris et al., 2006; Muse, Harris, & Field, 2003), job design (Champoux, 1992; Xie & Johns, 1995), person–environment fit (Edwards & Shipp, 2007), and the quality of work life (Rice, McFarlin, Hunt, & Near, 1985). More generally, developing propositions that describe various functional forms can be facilitated by intentionally considering alternatives to the usual linear relationship. As argued by Guion (1998, pp. 361-362):

> We are much too married to linear modeling. It is my guess that truly linear regression does not occur in nature unless we have restricted range, specifically restricted to the middle of a complete distribution.... I have been among those who have some dissatisfaction with an automatic assumption of linearity, but I do not think we need to abandon linear models universally. What is needed is an expansion of the options and perspectives of modeling by inviting the development and testing of nonlinear models as well.

Develop Contingent Predictions

Theories can be made more precise by developing contingent predictions. In general terms, the contingencies stipulated by a theory can be conceived as boundaries conditions that describe limits beyond which the theory does not apply (Bacharach, 1989; Dubin, 1978; Whetten, 1989). These boundary conditions can refer to populations, setting, time frames, and other circumstances under which the theory as a whole is deemed relevant. In more specific terms, contingencies refer to factors that influence the magnitude and form of the individual relationships predicted by the theory. These factors can be cast as moderator variables (Aiken & West, 1991; Jaccard, Turrisi, & Wan, 1990) and incorporated into propositions by stating how the proposed relationship between two constructs should vary across levels of the moderator variable. Whether treated as boundary conditions or moderator variables, contingencies built into theories increase precision by specifying when and how the relationships predicted by the theory should vary and whether the theory *in toto* is applicable in particular circumstances.

Contingencies strengthen theoretical precision by increasing the number of conditions that must be satisfied to infer support for the theory. These conditions become apparent when contingencies refer to moderator variables. To illustrate, consider the following moderated regression equation:

$$Y = b_0 + b_1X + b_2Z + b_3XZ + e. \tag{2}$$

In Equation 2, X and Y are independent and dependent variables whose relationship is described by a theoretical proposition, and Z is a moderator variable that influences the sign and magnitude of the relationship between X and Y. The impact of Z on the relationship between X and Y can be seen by rewriting Equation 2 in terms of simple slopes (Aiken & West, 1991):

$$Y = (b_0 + b_2Z) + (b_1 + b_3Z)X + e. \tag{3}$$

Equation 3 shows that Z influences the intercept and slope of the function that describes the relationship between X and Y. To illustrate, if the moderator variable of interest is gender, such that a theory predicts different relationships between X and Y for men and women, then Z can be conceived as a dummy variable coded 0 for men and 1 for women. With this specification, the equation for men can be written by setting $Z = 0$ in Equation 3, which yields:

$$Y = b_0 + b_1X + e. \tag{4}$$

Likewise, the equation for women is written by setting $Z = 1$, in which case Equation 3 becomes:

$$Y = (b_0 + b_2) + (b_1 + b_3)X + e. \tag{5}$$

Equations 4 and 5 show that incorporating gender as a moderator of a proposed relationship between X and Y effectively requires the theorist to specify two relationships, one for men and another for women. If the theory is only sufficiently strong to justify directional predictions, then the propositions concerning Equations 4 and 5 might state that the relationship between X and Y is positive for men and stronger (i.e., more positive) for women. The positive relationship for men would require that b_1 is greater than zero and the stronger relationship for women would stipulate that b_3 is greater than zero.[1] Although these predictions are merely directional, both must be satisfied to support the conditional proposition. Theoretical precision could be further enhanced by specifying range null hypotheses for each prediction, setting upper and lower limits for each effect, and introducing nonnil predictions as justified on conceptual and empirical grounds.

Contingent predictions can be developed in various ways. One approach is to draw from meta-analyses, which routinely document variation in effect sizes as evidence of moderation (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hunter & Schmidt, 2004; Lipsey & Wilson, 2000). In addition to indicating when moderation is likely, meta-analyses provide information about the magnitude of moderator effects, which can be used to calibrate moderator hypotheses. This use of meta-analysis to examine moderation was emphasized by Hall and Rosenthal (1991, p. 447), who added "If we want to know how well we are doing in the biological, psychological, and social sciences, an index that will serve us well is how far we have advanced in our understanding of the moderator variables of our field." This understanding is evidenced not only by documenting moderator effects in meta-analysis, but also by using those effects to inform the development of theoretical propositions that incorporate moderation. Contingent predictions can also result from systematically questioning whether the relationships predicted by a theory are universal. We believe this type of reasoning is often responsible for contingent predictions, perhaps to acknowledge that the phenomena explained by management theories have complex origins or to formalize the adage "it depends" that we often use to hedge our theoretical bets.

Summary

Theoretical precision can be increased by deriving propositions that narrow the set of outcomes that would be considered consistent with the theory. This type of refinement can be achieved by expanding the values that constitute evidence against the theory, as when null hypotheses are expressed as ranges rather than points and when upper and lower limits for parameter values are established. This refinement also results when propositions are stated as comparisons, such as predicting that a relationship is not only greater than zero but is also larger than another relationship. The outcomes taken as support for a theory are also narrowed when propositions are stated not as deviations from zero but as nonzero quantities supplemented by good-enough belts that specify ranges of allowable values (Serlin & Lapsley, 1985). Theories are also made more precise by specifying the functional forms of the relationships predicted by the theory and developing contingent predictions that identify factors that influence the form and magnitude of the relationships of the theory.

To clarify these methods for increasing theoretical precision, Table 1 illustrates how propositions that describe the relationship between two variables can be made more precise. For this illustration, the variables are labeled X and Y and refer to theoretical constructs, not observed measures, and the correlations and regression equations describing the relationships between X and Y are theoretical statements as opposed to empirically estimated expressions. As a point of departure, Table 1 begins with a standard directional prediction, which states that the relationship between X and Y is positive, or $r_{XY} > 0$. Expanding the null hypothesis would require that the relationship between X and Y exceeds a lower limit symbolized as r_L, which could represent the expected crud factor (Meehl, 1990a) or ambient noise (Lykken, 1968) for the relationship under consideration. Setting upper and lower limits would add that the relationship should not exceed some upper bound shown in Table 1 as r_U, which might represent the maximum correlation for which X and Y would be considered conceptually distinct. The proposed relationship can be stated as a comparison by predicting that Y will correlate more highly with X than with another variable W. As given in Table 1, this prediction is equivalent to the directional prediction $r_{XY} - r_{WY} > 0$, which can be further refined by setting upper and lower limits for difference between r_{XY} and r_{WY}. A nonnil proposition can be formulated by

Table 1. Illustrating Methods for Increasing the Precision of Theoretical Propositions

Method	Proposition	Formal statement
Directional prediction	X is positively related to Y	$r_{XY} > 0$
Expand the null hypothesis	X is positively related to Y with a correlation greater than the lower limit r_L	$r_{XY} > r_L$
Set upper and lower limits	X is positively related to Y with a correlation greater than the lower limit r_L and less than the upper limit r_U	$r_L < r_{XY} < r_U$
State predictions as comparisons	The relationship between X and Y is larger than the relationship between W and Y beyond the minimum threshold r_L	$r_{XY} > r_{WY}$
Develop nonnil predictions	X is positively related to Y with a correlation within the good-enough belt r_{GB} surrounding the correlation r_T	$r_T - .5r_{GB} < r_{XY} < r_T + .5r_{GB}$
Specify functional forms	X has a positive linear relationship with Y	$Y = b_0 + b_1X + b_2X^2 + e$ $b_1 > 0$ $b_2 \approx 0$
Develop contingent predictions	X has a positive relationship with Y that increases as the level of moderator Z increases	$Y = b_0 + b_1X + b_2Z + b_3XZ + e$ $b_1 + b_3Z > 0$ for all Z $b_3 > 0$

Note: X, Y, W, and Z refer to theoretical constructs. For methods that involve specifying functional forms and developing contingent predictions, the table includes theoretical expressions written as regression equations along with the pattern of coefficients that represent the stated proposition.

postulating that the correlation between X and Y will fall within a good-enough belt of width r_{GB} centered on the theoretical correlation r_T. To propose a functional form, a theoretical equation should be chosen that includes terms required to confirm or refute the function. In Table 1, the proposed function is positive and linear, which is expressed by stating that b_1, the coefficient on X, is greater than zero and that b_2, the coefficient on X^2, is approximately equal to zero.[2] The precision of these predictions can be enhanced by specifying upper and lower limits for b_1, stating that b_1 will fall within a good-enough belt of some nonzero value, or setting boundaries for a good-enough belt around zero that should encompass b_2. Finally, a contingent prediction is given in which X has a positive relationship with Y that increases as a function of Z. Drawing from Equation 3, the relationship between X and Y can be expressed as $b_1 + b_3Z$, which reinforces the notion that the slope relating X to Y varies as a function of Z. Predicting that X has a positive relationship with Y means that the expression $b_1 + b_3Z$ is positive for all theoretically plausible values of Z, and predicting that this relationship increases as a function of Z means that b_3 is positive. These two conditions are stated as directional predictions in Table 1, and their precision can be enhanced by specifying upper and lower boundaries or developing nonzero predictions surrounded by good-enough belts, as with the other directional predictions in Table 1.

Overcoming Barriers to Theoretical Precision

Theoretical precision is a lofty goal. Few researchers would deny the importance of pursuing theoretical precision, given that the alternative is to tolerate theories that yield vague predictions and slip through the clutches of rigorous empirical research. Nonetheless, making theories more precise is not an easy task. Some difficulties involve establishing a reasonable basis for justifying nonzero quantities, such as ranges around null hypotheses and predictions stated as nonnil values. Other difficulties are embedded in the broader enterprise of management research, where institutional forces and methodological practice have fostered an environment in which imprecise theories are tolerated. In this section, we discuss aspects of management research that present barriers to theoretical precision and suggest ways to overcome those barriers.

Focus Attention on Precision

One barrier to theoretical precision is the mere fact that precision is not emphasized in writings on theory development. Prescriptions for developing theory emphasize the importance of clearly defining constructs, describing relationships among constructs, explaining why these relationships exist, and specifying boundary conditions (Bacharach, 1989; Dubin, 1978; Sutton & Staw, 1995; Whetten, 1989). Beyond these fundamentals, we are encouraged to invoke disciplined imagination (Weick, 1989), incorporate multiple paradigms (Lewis & Grimes, 1999), draw from narratives and case studies (Eisenhardt, 1989; Pentland, 1999), resolve paradoxes (Poole & Van de Ven, 1989), and solve mysteries (Alvesson & Kärreman, 2007). These prescriptions are valuable, and following them can certainly facilitate theory development. However, these prescriptions do not push researchers to develop theories that are precise, such that their predictions are sufficiently bold to put the theory at risk of falsification. We tolerate theoretical imprecision not because we have little to say about theory development, but because what we say rarely mentions precision as something we should strive to attain (Gigerenzer, 1998).

The pursuit of precision invites us to expand the goals of theory development. Paying attention to precision does not mean we should place less emphasis on defining constructs or describing and explaining their relationships, nor does it mean we should abandon the creative devices and inductive procedures that inspire the theory development process. Rather, pursuing precision means that the goal of theorizing is not only to address the questions of what, how, why, who, when, and where (Dubin, 1978; Sutton & Staw, 1995; Whetten, 1989) but also the question of how much. The importance of this question was underscored by Tukey (1969, p. 86) in these terms:

> The physical sciences have learned much by storing up amounts, not just directions. If, for example, elasticity had been confined to “When you pull on it, it gets longer!” Hooke’s law, the elastic limit, plasticity, and many other important topics could not have appeared.

Answering questions of magnitude means that, when developing theory, we are not satisfied by merely saying that constructs will relate to one another. Instead, we strive to predict how much they will relate, test those predictions empirically, and refine them iteratively through a program of research that progressively strengthens theory (Runkel & McGrath, 1972). This process should attend to the functional form and contingencies of the relationships explained by theories, which contribute further to precision. In short, a simple but essential step toward theoretical precision is to keep the goal of precision clearly in focus as we develop, test, and refine management theories.

Challenge Norms of Theoretical Pluralism

Another barrier to theoretical precision is that making theories more precise puts them at greater risk of rejection, which runs counter to norms of theoretical pluralism that encourage us to embrace a potpourri of theoretical perspectives and sanctify theory development as an end in itself (Hambrick, 2007; McKinley, 2010). Theoretical pluralism can help advance knowledge, particularly at early stages of inquiry when researchers attempt to make sense of phenomena that are novel and complex. Theoretical pluralism also facilitates strong inference research that pits competing hypotheses against one another (Platt, 1964) and helps researchers identify alternative explanations for observations that are discrepant with prior theory (Lakatos, 1978; Meehl, 1978; Popper, 1959). However, when carried to excess, theoretical pluralism can encourage the development of theory for its own sake, such that theories are viewed not as means for guiding empirical research, but as ends in themselves. McKinley (2010, p. 48) describes this problem in the context of research on organization theory:

> In contemporary organization theory ... there has arguably been a displacement of ends in which new theory development has emerged as the ultimate end and the goal of consensus about the validity status of theory has become submerged ... the empirical activities of instrumental standardization, empirical testing of existing theories, and replication of those tests have been relegated to a position of second priority relative to new theory development.

Treating theory development as an end has resulted in a proliferation of untested theories (Eden, 2004; Hambrick, 2007; Kacmar & Whitfield, 2000; McKinley, 2010). Eden (2004, p. 171) argued that this situation is aggravated by the policies of leading journals, such as the *Academy of Management Journal* (*AMJ*), which require authors to make novel theoretical contributions with each submission:

> under *AMJ*'s current norms, the journal would reject your work for not providing a novel theoretical contribution. This policy ... encourages the development of one-time minitheories that serve as platforms for specific submissions but are never revisited. This policy has left a trail of one-shot theories that contribute little to the field's cumulative scientific endeavor.

Pfeffer (1993, p. 616) offered an alternative perspective, arguing that the proliferation of theories results from weak paradigms that characterize management research:

> I believe that the field encourages the development and advancement of differences and separate agendas rather than attempts at integration or resolution. More than 10 years ago, I (Pfeffer, 1982, p. 1) argued that "the domain of organization theory is coming to resemble more of a weed patch than a well-tended garden. Theories ... proliferate along with measures, terms, concepts, and research paradigms. It is often difficult to discern in what direction knowledge of organizations is progressing." The situation has not changed.

These forces contribute an environment in which more theory is considered inherently good and the development of new theory takes precedent over the investigation, let alone the rejection, of existing theory.

To promote theoretical precision, the prevailing norms of theoretical pluralism should be replaced with norms of theoretical refinement, whereby researchers are encouraged to rigorously evaluate theories, modify or discard propositions that fail to survive empirical tests, and develop new theory when existing theories prove untenable. These norms promote the belief that putting weak theories out of their misery is just as noble as creating new theories. To pursue Pfeffer's (1982) analogy, we should refine the theoretical landscape by pulling conceptual weeds and thinning out theories that are manifestly weaker than their competitors. Norms of theoretical refinement can be promoted by supporting the publication of empirical work that tests existing theories and replications that expose theories to repeated scrutiny (Eden, 2004; Hambrick, 2007; Hubbard, Vetter, & Little, 1998; McKinley, 2010; Neuliep & Crandall 1991). Norms of theoretical refinement mesh well with the goal of theoretical precision, because refinement narrows the range of possibilities allowed by a theory, and theories stated in precise terms are more subject to evaluation and refinement than theories laden with imprecision.

Embrace Negative Results

A third barrier to theoretical precision arises from bias against negative results in empirical research. This bias has been documented by investigations of the review process, which show that studies with findings that run counter to a priori hypotheses are less likely to receive favorable evaluations from reviewers and editors (Hubbard & Armstrong, 1997; Mahoney, 1985; Pfeffer, 2007; Sterling, Rosenbaum, & Weinkam, 1995). This tendency is compounded by the behavior of authors themselves, who are prone to withhold results that fail to support hypotheses (Rotton, Foos, Vanmeek, & Levitt, 1995), leading to what has been termed the "file drawer problem" (Rosenthal, 1979). The bias against negative results runs counter to the pursuit of theoretical precision, which encourages us to generate specific hypotheses that risk rejection. If rejecting hypotheses is considered undesirable by gatekeepers of the publication process and authors attempting to navigate this process, then researchers would be discouraged from creating precise theories with bold, risky hypotheses.

Overcoming this barrier would require management researchers to reframe how they think about the goals of empirical research. Our interactions with students and colleagues have led us to believe that many researchers implicitly or explicitly assume that the goal of empirical research is to demonstrate support for theories. When commenting on research in organization theory, McKinley's (2010, p. 56) observed:

> These theories are engaging, and even sometimes exciting, although the empirical support offered for them often seems to be directed more toward legitimizing the theory than toward facilitating additional empirical testing and producing a discipline-wide consensus about whether or not a given theory is valid.

In line with this observation, we have heard many researchers lament that a study "did not work" when it failed to support the proposed hypotheses. We submit that whether a study has worked should not be judged by the outcomes of hypothesis tests but instead by whether the study was well conceived, designed, and executed, and whether it embodied sound methodological practice (e.g., used measures with strong reliability and validity and drew a sample large enough to yield adequate statistical power). If a study is conducted well, then it has "worked," regardless of whether the hypotheses examined by the study are corroborated or rejected (Edwards, 2008). In this sense, empirical research is akin to going down alleys to see if they are blind, and the excursion has served its purpose regardless of whether it leads to a dead end or an open door.[3]

Leverage Empirical Results

Some techniques for increasing theoretical precision draw on results from empirical research to specify hypotheses. This practice might seem inherently atheoretical because it allows empirical results to influence what we subsequently predict (Sutton & Staw, 1995). Empirically driven research is generally frowned upon, and those who conduct such research are sometimes chided as conducting fishing expeditions, dredging data, or resorting to dust-bowl empiricism. Even when empirical results are driven by a priori hypotheses, researchers rarely use their results for anything other than determining whether hypotheses are supported, paying little heed to the sizes of the obtained effects (Kirk, 1996). The tendency to disregard effect sizes is also evidenced by the ways researchers use published meta-analyses, in that most researchers cite meta-analyses to establish that effects exist rather than to describe the sizes of the effects, let alone use them to calibrate hypotheses for subsequent studies (Carlson & Ji, 2009).

To increase our willingness to leverage empirical results, we should place greater value on empiricism and extract more information from the accumulated findings at our disposal. We are not suggesting that researchers should draw generalizations from individual studies or treat every empirical finding as theoretically meaningful. Rather, we encourage researchers to draw from empirical regularities, or stylized facts (Helfat, 2007), which are replicable and make sense from a conceptual standpoint. These regularities should be used to establish thresholds for range null hypothesis, good-enough belts for nonnull hypotheses, parameters that describe functional forms, and other factors that bring precision to hypotheses and the theories from which they are derived. Moreover, the information drawn from empirical results should go beyond whether an effect was found to include the magnitude of the effect and the surrounding confidence interval. If we become more accustomed to interpreting empirical results in terms of their magnitude, then we are likely to become more comfortable framing propositions and hypotheses in similar terms. In this manner, conceptual reasoning can indicate whether and why an effect exists, and empirical results that are replicated by methodologically sound studies can help us calibrate the size of the effect.

Develop Meaningful Metrics

Theoretical precision is also inhibited by the arbitrary metrics of many measure used in management research. Metrics are arbitrary when the scores yielded by a measure have no inherent mapping onto the underlying construct being measured (Blanton & Jaccard, 2006; Sechrest, McKnight, & McKnight, 1996). For instance, a score of 5 on a job satisfaction scale ranging from 1 to 7 does not itself indicate the particular level of satisfaction experienced by the respondent. It is probably safe to assume that higher scores indicate greater job satisfaction, but the amount of satisfaction corresponding to any given score is not conveyed by the score itself. In contrast, nonarbitrary metrics are inherently meaningful and can be directly interpreted, as when firm profit is measured in dollars, absenteeism is measured as days of missed work, and job performance is measured as the number of units produced.

Arbitrary metrics impede theoretical precision because statements about the size of the relationship between variables depend on the metrics of the variables involved. If the variables have arbitrary metrics, then quantities used to express the size of the relationship between the variables (e.g., regression coefficients) are likewise arbitrary. As noted by Tukey (1969, p. 89):

> Given two perfectly meaningless variables, one is reminded of their meaninglessness when a regression coefficient is given, since one wonders how to interpret its value. . . . Being so uninterested in our variables that we do not care about their units can hardly be desirable.

The arbitrariness of metrics used to measure variables and the parameters that describe the relationships between variables inhibit the placement of boundaries around null hypotheses, the location and range of nonnil hypotheses, and other quantities used to state predictions in precise terms. Although the ambiguities of arbitrary metrics are usually ascribed to measures, they also apply to constructs, which are usually characterized as having no scales of their own. When constructs and their measures both have arbitrary metrics, it is difficult to establish the foothold needed to take the initial steps toward making metrics more meaningful.

Metrics can be made less arbitrary, and hence more meaningful, using various tactics. For instance, the response options of measurement scales can be assigned specific anchors that give meaning to each numeric value of the scale (Bass, Cascio, & O'Connor, 1974; Schriesheim & Castro, 1996). This approach underlies behaviorally anchored rating scales (Kingstrom & Bass, 1981; Smith & Kendall, 1963), which have scale points labeled with concrete behaviors that describe varying levels of performance effectiveness. Alternately, substantive outcomes associated with each level of a rating scale can be identified (Blanton & Jaccard, 2006; Sechrest et al., 1996). For instance, the meaning of scores on a measure of turnover intent could be clarified by documenting the proportion of employee turnover associated with each score. Metrics can also be made more meaningful by asking subject matter experts to give their interpretation of each response option provided by a measure. This approach has been recommended for measures used in psychotherapy, with the goal of developing evidenced-based treatments that produce patient outcomes judged to be clinically important (Kazdin, 2006; Sechrest et al., 1996). Although these tactics focus on reducing the arbitrariness of metrics for measures, they also make the metrics of constructs less arbitrary, given that scales for constructs can be established by specifying the function linking a construct to a referent indicator, as is common practice in structural equation modeling (Bollen, 1989).

We make no pretense that the arbitrariness of metrics used in management research can be completely eliminated, producing scales akin to those used to measure height, weight, or temperature. However, we should not revert to the opposite extreme, succumbing to the belief that the metrics we use mean nothing at all. When respondents choose options on a scale, they are assigning meaning to those options and calibrating them to the levels of the belief, attitude, or judgment we want to assess (Sudman, Bradburn, & Schwarz, 1996). As such, respondents do not treat our metrics as arbitrary, and therefore we should not consider them arbitrary simply because it might be unclear how the obtained scores map onto levels of the construct of interest. Rather, we should seek to understand how respondents interpret the metrics we provide and work toward consensus regarding the meaning of terms such as "small" and "large," "low" and "high," and other labels that describe the levels of constructs and measures.

Move Beyond Null Hypothesis Significance Testing

A final barrier to precision is the persistence of null hypothesis significance testing in management research. Although null hypothesis significance testing has been criticized for decades (Bakan, 1966; Carver, 1978; Cohen, 1994; Krueger, 2001; Nickerson, 2000; Rozeboom, 1960), its use continues throughout the social sciences. As noted at the outset of this article, null hypothesis significance testing is likely to continue in management research until theories yield predictions other than deviations from zero. At the same time, the ongoing acceptance of null hypothesis significance testing undermines the development of precise theories. As put by Gigerenzer (1998, p. 200):

> As long as there is an institutionalized methodology that does not encourage researchers to specify their hypotheses, there is little incentive to think hard and develop theories from which such hypotheses could be derived.

The manner in which null hypothesis significance testing impedes the development of strong theory in psychology was underscored by Dar (1987, p. 149) as follows:

> null hypothesis testing and related practices seem to replace good theory building ... instead of demanding a high level of logical consistency, explanatory power, and accurate predictions from their theories, psychology researchers are trained to feel satisfied when a relevant statistic (e.g., a correlation coefficient) is statistically different from zero.... When passing null hypothesis tests becomes the criterion for successful predictions, as well as for journal publications, there is no pressure on the psychology researcher to build a solid, accurate theory; all he or she is required to do, it seems, is produce "statistically significant" results.

This stalemate characterizes management research, such that our continued tolerance of null hypothesis significance testing distracts us from the value of strong theory. If null hypothesis significance testing is our analytical stock in trade, why bother developing theories that predict anything other than deviations from null values?

The value of increasing theoretical precision would come into sharp relief if we took to heart the criticisms of null hypothesis significance testing. Although null hypothesis significance testing can be useful for certain purposes (Abelson, 1997; Cortina & Dunlap, 1997; Frick, 1996; Wainer, 1999), its ritualistic application obscures the value of developing theories that predict anything other than what null hypothesis significance tests bring into focus, which is whether a parameter differs from zero. The persistence of null hypothesis significance testing exemplifies the inertia of methodological practice, which can be slow to change even in the face of mounting criticism (Cohen, 1994; Falk & Greenbaum, 1995; Krueger, 2001). Breaking the habit of null hypothesis significance testing would help researchers see value in theoretical precision and, at the same time, precise theories can yield predictions that would prompt researchers to apply alternatives to null hypothesis significance testing.

Summary and Conclusion

Increasing theoretical precision is a challenging pursuit. Doing so requires management researchers to push theory development beyond current practice in which propositions take the form of directional predictions. Such propositions bear little risk of falsification when tested with methodologically strong studies, and the corresponding theories have little at stake when evaluated against empirical evidence. This state of affairs undermines scientific progress in management research and impedes our ability to make definitive statements that shape policy and demonstrate that we have learned something definitive. We can do better, and the starting point of this endeavor traces back to the precision of the theories we develop, test, and refine. Making theories more precise requires us to tackle challenges that include calibrating effect sizes, specifying functional forms, developing meaningful metrics, and confronting norms that value theoretical pluralism, hypothesis confirmation, and relying on the familiar hammer of null hypothesis significance testing. These challenges are nontrivial, but they can and should be addressed in the interest of developing theories that predict the presence of something rather than the absence of nothing in management research.

Notes

1. In most cases, if b_1 and b_3 are both statistically greater than zero, then the sum $b_1 + b_3$ will also be statistically greater than zero. Therefore, establishing that the relationship between X and Y is positive for men and stronger (i.e., more positive) for women implies that the relationship is positive for women. However, the opposite is not true, such that when either b_1 or b_3 does not statistically differ from zero, the sum $b_1 + b_3$ may or may not statistically differ from zero. Thus, to fully evaluate contingent predictions for dichotomous moderators such as gender, it is useful to statistically test b_1, b_3, and $b_1 + b_3$. This reasoning can be extended to categorical moderators with more than two levels (Aguinis, 2003) and can also be extended to continuous moderators, in which case the expressions for simple slopes such as those in Equations 4 and 5 can be written as $b_1 + b_3Z_L$ and $b_1 + b_3Z_H$, where Z_L and Z_H are low and high levels of Z, respectively, chosen for comparison by the researcher (Aiken & West, 1991).
2. Establishing that b_2 does not meaningfully deviate from zero indicates that the function relating X to Y is not quadratic. Additional higher order terms, such as X^3, X^4, and so forth, could be considered to examine other departures from linearity, acknowledging that the risk of capitalizing on chance increases as more terms are added to the equation (Pedhazur, 1997).
3. We are indebted to Cary L. Cooper for this metaphor of the research process.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, *8*, 12-15.

Adner, R., Polos, L., Ryall, M., & Sorenson, O. (2009). The case for formal theory. *Academy of Management Review*, *34*, 201-208.

Aguinis, H. (2003). *Regression analysis for categorical moderators.* New York, NY: Guilford.

Aiken, L. A., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review*, *32*, 1265-1281.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37-47.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two step approach. *Psychological Bulletin*, *103*, 411-423.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*, 496-515.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*, 421-458.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423-437.

Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, *59*, 313-320.

Binder, A. (1963). Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, *70*, 107-115.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27-41.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley.

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

Carlson, K. D., & Ji, F. X. (2009). *Citing, reporting, interpreting, and building on metaanalytic findings*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Education Review*, *48*, 378-399.

Carver, R. P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, *61*, 287-292.

Champoux, J. E. (1992). A multivariate analysis of curvilinear relationships among job scope, work context satisfactions, and affective outcomes. *Human Relations*, *45*, 87-111.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*, 997-1003.

Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods*, *2*, 161-172.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147-167.

Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psychologists. *American Psychologist*, *42*, 145-151.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing theory through simulation methods. *Academy of Management Review*, *32*, 480-499.

Diener, E., Ng, W., & Tov, W. (2009). Balance in life and declining marginal utility of diverse resources. *Applied Research in Quality of Life*, *3*, 277-291.

Dubin, R. (1978). *Theory building.* New York, NY: Free Press.

Eden, D. (2004). Reflections on the *AMJ* Associate Editor Role. *Academy of Management Journal*, *47*, 167-173.

Edwards, J. R. (2008). To prosper organizational psychology should ... overcome methodological barriers to progress. *Journal of Organizational Behavior*, *29*, 469-491.

Edwards, J. R., & Shipp, A. J. (2007). The relationship between person-environment fit and outcomes: An integrative theoretical framework. In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit* (pp. 209-258). San Francisco, CA: Jossey-Bass.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*, 532-550.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, *5*, 75-98.

Ferris, G. R., Bowen, M. G., Treadway, D. C., Hochwarter, W. A., Hall, A. T., & & Perrewé, P. L (2006). The assumed linearity of organizational phenomena: Implications for occupational stress and well-being. In P. L. Perrewé & D. C. Ganster (Eds.), *Research in occupational stress and well-being* (Vol. 5, pp. 205-232). Oxford, UK: Elsevier Science Ltd.

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying nonzero effect null hypotheses. *Journal of Applied Psychology*, *70*, 215-218.

Freese, L. (1980). Formal theorizing. *Annual Review of Sociology*, *6*, 187-212.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, *1*, 379-390.

Gigerenzer, G. (1998). Surrogates for theories. *Theory and Psychology*, *8*, 195-204.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587-606.

Gillett, R. (1994). Post hoc power analysis. *Journal of Applied Psychology*, *79*, 783-785.

Guion, R. M. (1998). Some virtues of dissatisfaction in the science and practice of personnel selection. *Human Resource Management Review*, *8*, 351-365.

Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs*, *58*, 437-448.

Hambrick, D. C. (2007). The field of management‘s devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*, 1346-1352.

Harrison, J. R., Lin, Z., Carroll, G. R., & Carley, K. M. (2007). Simulation modeling in organizational and management research. *Academy of Management Review*, *32*, 1229-1245.

Helfat, C. E. (2007). Stylized facts, empirical research and theory development in management. *Strategic Organization*, *5*, 185-192.

Hodges, J., & Lehmann, E. (1954). Testing the approximate validity of statistical hypotheses. *Journal of the Royal Statistical Society (B)*, *16*, 261-268.

Hubbard, R., & Armstrong, J. S. (1997). Publication bias against null results. *Psychological Reports*, *80*, 337-338.

Hubbard, R., Vetter, D. E., & Little, E. L. (1998). Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal*, *19*, 243-254.

Hulin, C. L. & Ilgen, D. R. (Eds.). (2000). *Computational modeling of behavior in organizations.* Washington, DC: American Psychological Association.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Johns, G. (1991). Substantive and methodological constraints on behavior and attitudes in organizational research. *Organizational Behavior and Human Decision Process*, *49*, 80-104.

Kacmar, K. M., & Whitfield, J. M. (2000). An additional rating method for journal articles in the field of management. *Organizational Research Methods*, *3*, 392-406.

Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist*, *61*, 42-49.

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (bars) and other rating formats. *Personnel Psychology*, *34*, 263-289.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746-759.

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist*, *56*, 16-26.

Lakatos, I. (1978). Falsification and the methodology of scientific research programs. In J. Worrall & G. Currie (Eds.), *The methodology of scientific research programs: Imre Lakatos philosophical papers* (Vol. 1). Cambridge, England: Cambridge University Press.

Land, K. C. (1971). Formal theory. *Sociological Methodology*, *3*, 175-220.

Lane, R. E. (2000). Diminishing returns to income, companionship—and happiness. *Journal of Happiness Studies*, *1*, 103-119.

Lewis, M. W., & Grimes, A. J. (1999). Metatriangulation: Building theory from multiple paradigms. *Academy of Management Review*, *24*, 672-690.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Thousand Oaks, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis.* Newbury Park, CA: Sage.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151-159.

MacCallum, R., Roznowski, M., Mar, C. M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate Behavioral Research*, *29*, 1-32.

Mahoney, M. J. (1985). Open exchange and epistemic progress. *American Psychologist*, *40*, 29-39.

McKinley, W. (2010). Organizational theory development: Displacement of ends? *Organization Studies*, *31*, 47-68.

McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York, NY: John Wiley.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103-115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806-834.

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*, 108-141.

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244.

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 391-423). Hillsdale, NJ: Lawrence Erlbaum.

Mosier, C. I. (1951). Symposium: The need and means of cross-validation. *Educational and Psychological Measurement*, *11*, 5-11.

Mulaik, S. A., Raju, N. S., & Harshman, R. A. (1997). There is a time and place for significance testing. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 65-116). Hillsdale, NJ: Erlbaum.

Murphy, K. R. (1990). If the null hypothesis is impossible, why test it? *American Psychologist*, *45*, 403-404.

Murphy, K. R., Myors, B., & Wolach, A. (2009). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Muse, L. A., Harris, S. G., & Field, H. S. (2003). Has the inverted-U theory of stress and job performance had a fair test? *Human Performance*, *16*, 349-364.

Neuliep, J. W., & Crandall, R. (1991). Editorial bias against replication research. In J. W. Neuliep (Ed.), *Replication research in the social sciences* (pp. 85-90). Newbury Park, CA: Sage.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*, 241-301.

Northcraft, G. B., Preston, J. N., Neale, M. A., Kim, P. H., & Thomas-Hunt, M. C. (1998). Non-linear preference functions and negotiated outcomes. *Organizational Behavior and Human Decision Processes*, *73*, 54-75.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York, NY: Holt.

Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review*, *24*, 711-724.

Pfeffer, I. (1982). *Organizations and organization theory.* Marshfield, MA: Pitman.

Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, *18*, 599-620.

Pfeffer, J. (2007). A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal*, *50*, 1334-1345.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347-353.

Poole, M. S., & Van de Ven, A. H. (1989). Using paradox to build management and organization theories. *Academy of Management Review*, *14*, 562-578.

Popper, K. R. (1959). *The logic of scientific discovery.* New York, NY: Basic Books.

Rice, R. W., McFarlin, D. B., Hunt, R. G., & Near, J. P. (1985). Organizational work and the perceived quality of life: Toward a conceptual model. *Academy of Management Review*, *10*, 296-310.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1-12.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Rotton, J., Foos, P. W., Vanmeek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behavior and Personality*, *10*, 1-13.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*, 416-428.

Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method.* New York, NY: Holt.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173-1181.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, *10*, 1-22.

Schriesheim, C. A., & Castro, S. L. (1996). Referent effects in the magnitude estimation scaling of frequency expressions for response anchor sets: An empirical investigation. *Educational and Psychological Measurement*, *56*, 557-569.

Sechrest, L., McKnight, P., & McKnight, K. (1996). Calibration of measures in psychotherapy outcome studies. *American Psychologist*, *51*, 1065-1071.

Serlin, R. C. (1993). Confidence intervals and the scientific method: A case for Holm on the range. *Journal of Experimental Education*, *61*, 350-360.

Serlin, R. C., & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist*, *40*, 73-83.

Smith, P. C., & Kendall, L. (1963). Retranslation of expectation: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, *47*, 149-155.

Snee, R. D. (1977). Validation of regression models: Methods and examples. *Technometrics*, *19*, 415-428.

Spencer, B. (1995). Correlations, sample size, and practical significance: A comparison of selected psychological and medical investigations. *Journal of Psychology*, *129*, 469-475.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*, 108-112.

Stimson, J. A., Carmines, E. G., & Zeller, R. A. (1978). Interpreting polynomial regression. *Sociological Methods & Research*, *6*, 515-524.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco, CA: Jossey-Bass.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371-384.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminancy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371-386.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*, 83-91.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science, 6*, 100-116.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions, and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 165-191). New York, NY: Routledge.

Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods*, *4*, 212-213.

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*, 516-531.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*, 490-495.

Xie, J. L., & Johns, G. (1995). Job scope and stress: Can job scope be too high? *Academy of Management Journal*, *38*, 1288-1309.

Bios

Jeffrey R. Edwards is the Belk Distinguished Professor of Organizational Behavior at the Kenan-Flagler Business School, University of North Carolina. He is past editor of *Organizational Behavior and Human Decision Processes*, past chair of the Research Methods Division of the Academy of Management, and a fellow of the Academy of Management, the American Psychological Association, the Society for Industrial and Organizational Psychology, and the Center for the Advancement of Research Methods and Analysis. His methodological research addresses difference scores, polynomial regression, structural equation modeling, construct validity, and the development and evaluation of theory.

James W. Berry is a PhD Candidate in Organizational Behavior at the Kenan-Flagler Business School, University of North Carolina. His research focuses on creativity and how new knowledge is generated, evaluated, and ultimately accepted in organizations and disseminated throughout academic communities within management research. His creativity research has been published in the *Academy of Management Journal* and has been presented at national meetings for the Academy of Management and the Society for Industrial and Organizational Psychology.