Establishing the Usefulness of Strategic Management Research:

On Inverted Lewinians and Naked Strategy Scholars

Craig J. Russell

University of Oklahoma

I would like to thank my colleague Shaila Miranda for her helpful comments throughout and special assistance in creating Figure 1, though all flaws and shortcomings in this chapter remain my responsibility.

Earlier this year I had a conversation with Don Bergh that led to his kind invitation to write this chapter. A little background will help put that conversation and this chapter in context. I am not a strategic management scholar - my field of specialization is human resource management. HRM topics of interest to me required that I learn a number of research methods introduced to management literatures after I left graduate school in 1982, e.g., structural equation modeling (Russell, 1985), meta-analysis (Russell, Settoon, McGrath, Blanton, Kidwell, Lohrke, Scifries, & Danforth, 1994), hierarchical linear models (Russell, 2001), bootstrapping (Russell & Dean, 2000), and most recently item response theory. Collaboration with colleagues in other management sub-disciplines (e.g., strategic management, international management, MIS) came as a direct result of these interests. Most relevant to this chapter were meta-analyses of the strategic groups and cultural distance literatures (Ketchen, Combs, Russell, Shook, Dean, Runge, Lohrke, Naumann, Haptonstahl, Baker, Beckstein, Handler, Honig, & Lamoreaux, 1997; Tihanyi, Griffeth, & Russell, 2004). I continue to peruse the strategic groups and cultural distance literatures as an ad hoc reviewer for a number of management journals due to my involvement with these meta-analyses.

As a result of having read a great deal of original primary research in the strategic management literature for these meta-analyses, I mentioned to Don that I had been surprised by the number of basic research methods problems I encountered. To be sure, meta-analyses I have collaborated on in all literatures revealed a small percentage of what I have come to call "whoops" errors, named after the typical response received when I asked original authors for clarification of some curious or incongruous statistic reported in their article (e.g., degrees of freedom that don't add correctly, effect sizes reported in tables that differ from those reported in the text, and other "housekeeping" kinds of mistakes). A small number of truly scary methodological errors were also revealed in the process of meta-analyzing these literatures. For example, in the final stages of writing the last draft of Ketchen et al. (1997) after it had been accepted, a coauthor brought a doctoral dissertation to my attention that empirically examined strategic group evidence in a sample of over 100,000 "organizations." Only ~ 27,000 organizations contributed to the 40 effect sizes used in our meta-analysis. Inclusion of this one study would have overwhelmed the 40 effect sizes, changing virtually every conclusion drawn and drastically modifying the manuscript! In reading the dissertation I was stunned to find the author's sample came from approximately 13 years of CRISP tapes in which he (for example) had counted General Motors' data for 1980 through 1992 as information on 13 different and independent organizations. This violated fundamental assumptions required of empirical procedures used to create "strategic groups" in this literature (and used by the dissertation author), making all "groups," inferential statistics calculated from these "groups," and conclusions drawn about these "groups" spurious in this dissertation – its results were not included in the Ketchen et al. (1997) meta-analysis.[1]

I will not address these more conventional statistical issues, but instead focus on the way strategy scholars determine whether their theories or models are any good. Without having obtained an exact count, I came away with two dominant methodological concerns from the majority of primary research articles[2] I read in the Academy of Management Journal, Strategic Management Journal, Journal of International Business Studies, Academy of Management and Strategic Management Society meetings, and other outlets over the last 25-30 years. The major goal of this chapter is to describe these concerns and how strategy researchers might address it to

---

[1] The positive way to frame this dissertation is that, if it were ever submitted, it never made it through the referee process to be presented at a national conference or published in a scholarly journal.

[2] Secondary research uses results from primary research as its data (e.g., meta-analysis). Primary research evaluates hypotheses and research questions by drawing inferences from investigations of actual samples of the phenomena.

enhance theory development and performance prediction. I am very aware of the fact that the concerns I raise below occur (unfortunately) in HRM and other more "micro" management arenas, though I would contend with meaningfully lower frequencies. Regardless, the concerns raised below must be addressed if management research of all genres is to make real contributions to theory development and practice.

The first concern focused on here stems from the inverse of a view widely attributed to Kurt Lewin, i.e., that there is nothing as useful as a good theory. While I strongly agree with this sentiment, I would go further in applied arenas (e.g., business administration) and argue the inverse is also true, i.e., that a theory is not very good unless it is useful.[3] This begs the question of what constitutes "usefulness" in management theory. Note, this is not the "rigor vs. relevance" issue raised so often as of late (e.g., Gulati, 2007). It is instead simply a focus on how to demonstrate relevance, or "usefulness," of the rigorous research being reported. Hence, I will focus on a simple misinterpretation that occurs with alarming frequency. Specifically, all too often authors report and interpret coefficients of determination (i.e., $r_{xy}^2\ or\ R_{y-x_1x_2}^2$, or their parallels in logit, probit, feasible generalized least squares, and other alternatives used when OLS assumptions are not met) as metrics of how "good" a theory or model is in its prediction of some criteria of interest. We have all seen authors conclude Model A is clearly better than Model B because it explains more variance in some criterion Y. I will demonstrate why this is not an appropriate metric with which to judge a model's usefulness and discuss an alternative metric and its implications. In doing so, I will assume for purposes of this discussion that the only

[3] Lewin (1942) stated that there was nothing as practical as a good theory. Clearly, in non-applied arenas theory is also useful when it serves as a bridge to better theory.

relevant metrics of usefulness or "value" are economic, i.e., involve dollars (e.g. sales, profit, EBITDA, etc.).[4]

The second concern is with the near universal use of unwarranted causal language in interpreting results from strategic management research. After touching on the classic "correlation is not causation" observation, I will suggest explicit ways and provide examples of how to conduct strong tests of hypothesized causal strategic relationships. These two concerns – r_{xy}^2 as an inappropriate metric of a theory's usefulness and unjustified causal inferences – are two major reasons why Academy of Management and Strategic Planning Society research presentations are not overly subscribed by CEO's and Strategic Management V.P.'s. We need to assess theory quality using standards relevant to actual business executives (not just statisticians) and have more substance behind our causal inferences.

Concern I: The Brogdon-Cronbach-Gleser Model

(r_{xy} *vs.* r_{xy}^2 *or* $R_{y-x_1x_2\ldots x_k}$ *vs.* $R_{y-x_1x_2\ldots x_k}^2$)

Firms and individuals budget or account for dollars, <u>not</u> standardized dollars, squared dollars, squared deviations from mean dollars, or percentage of squared deviations from mean dollars – my checking account reports my balance in dollars. In contrast, we have all seen a model dismissed because it "only explained 9% of the variance." However, the Brogden-Cronbach-Gleser (BCG) model clearly shows that r_{xy} (or $R_{y-x_1\,x_2\ldots\,x_k}$) is linearly related to a model's dollar utility to the firm, <u>not</u> r_{xy}^2 or $R_{y-x_1x_2\ldots\,x_k}^2$. In other words, when r_{xy} (or $R_{y-x_1\,x_2\ldots\,x_k}$) doubles for a strategic management model designed to predict profit ($Y_\$$), then the predicted dollar value added to the firm doubles (e.g., when r_{xy} = .30 and $R_{y-x_1x_2}$ = .60, the

[4] I am aware of other criteria of import in organizational settings (e.g., "green" issues, ethical considerations, etc.), just as I am aware of the various types of non-profit organizations in existence. The methods described here may be applied to predicting outcomes other than dollar value that may be salient in such settings if quantitative measures of these outcomes were available.

addition of X_2 to the model has increased expected dollar value added to the firm by a factor of 2). Hence, a model that explains only 9% of the variance in $Y_\$$ in fact explains 30% of the dollar utility available to be explained in $Y_\$$, even though tests of the null hypothesis H_0: $r_{xy} = 0$ and H_0: $R_{y-x_1x_2\ldots x_k} = 0$ will yield mathematically identical outcomes to tests of H_0: $r^2_{xy} = 0$ and H_0: $R^2_{y-x_1x_2\ldots x_k} = 0$. Not surprisingly, I rarely see the BCG model cited in the scholarly management literature, and never see it cited by strategic management scholars. So, I will first demonstrate how the Brogden-Cronbach-Gleser (BCG) model was originally developed to show how personnel selection systems add value to firms, though it also characterizes how the dollar impact of any organizational intervention can be estimated, be it strategic, entrepreneurial, HR-related, etc. I will then make some minor adjustments to show how the model can be applied to more macro, strategic research arenas as well as some of the more interesting implications that are seldom fully appreciated in the current management literature. I will conclude this section with an example of how the BCG model might be applied to a recent strategic management study published in a recent issue of the Academy of Management Journal.

Brogden-Cronbach-Gleser. Brogden (1949) started with the following OLS regression model:

$$\hat{y}_\$ = b_0 + b_1 x_1$$

Equation 1

Three parameters estimated from sample data for simple OLS regression models are b_0, b_1, and the Pearson product moment correlation (r_{xy}). If we don't know yet how well someone is going to perform on a job (which we can't know before s/he is hired), then one estimate of how s/he might perform would be the $\hat{Y}_\$$ value obtained from plugging the applicant's personnel selection test score X_1 into **Error! Reference source not found.**. Ordinary least squares regression analyses give us the formula for the "best" fitting straight line (i.e., **Error! Reference source not found.**), where "best" means the formula for the straight line $\hat{y}_{\$} = b_0 + b_1 x_1$ that minimizes the sum of all squared prediction errors ($\sum_{i=1}^{N}(y_{\$_i} - \hat{y}_{\$_i})^2$) across people in the sample.

Let's assume the dependent measure is already in dollar terms (e.g., store profit, sales volume, etc., when selecting retail store managers). Brogden (1949) derived his model by first standardizing the predictor variable X, i.e., he standardized applicants' personnel selection test scores to create . . .

$$\hat{y}_{\$} = b_0 + b_1 z_i$$

Equation 2

Note, Brogden did not standardize $Y_{\$}$, as scholars and practitioners are interested in predicting dollars, not standardized dollars, or sums of squared deviations between predicted dollars and actual dollars.

Some final substitutions modify Equation 2 to show the financial impact expected from use of the personnel selection test in screening a group of applicants. Brogden started by taking the expected value of Equation 2 . . .

$$E(\hat{y}_{\$}) = E(b_0) + E(b_1)E(z_s)$$

$$\bar{y}_{\$} = E(b_0) + E(b_1)\bar{z}_s$$

Equation 3

Where $\bar{z}_s$ is the average standardized personnel test score for those applicants actually selected by the firm. When no selection system is used (i.e., if applicants had been chosen at random), $\bar{z}_s$ is expected to be the same as the average of z scores for all applicants, or $\bar{z}_s = 0$. When $\bar{z}_s = 0$ then $E(b_1)\bar{z}_s = 0$ too, and the remainder - $E(b_0)$ - will be the average dollar performance of individuals selected at random from the applicant pool. Using $\mu_\$$ as the symbol for expected or average dollar performance for everyone in the applicant pool, we can substitute $\mu_\$$ for $E(b_0)$ in Equation 3 . . .

$$\bar{y}_\$ = \mu_\$ + E(b_1)\bar{z}_s$$

Equation 4

Finally, the expected value of b_1 can be estimated directly from a sample obtained in a criterion-related validity study, though it is often useful to substitute for $E(b_1)$. Specifically, the sample regression coefficient or slope estimate is also defined as follows . . .

$$b_1 = r_{xy}\left(\frac{SD_y}{SD_x}\right)$$

Equation 5

where:

> r_{xy} = the simple Pearson product moment correlation between test scores on the personnel selection test x and the measure of job performance y.
>
> SD_y = the standard deviation of job performance measured in dollars
>
> SD_x = the standard deviation of all applicant's test score performance

However, recall applicant test scores were standardized in Equation 3 to create the z variable used in Equation 4. So, instead of $b_1 = r_{xy}\left(\frac{SD_y}{SD_x}\right)$, b_1 becomes $b_1 = r_{xy}\left(\frac{SD_y}{SD_z}\right)$. As the standard deviation of z scores is $SD_z = 1.0$, substituting 1 for SD_z, Equation 5 becomes $b_1 = r_{xy}SD_y$. So, substituting $\mu_\$$ for $E(b_0)$ and $r_{xy}SD_y$ for $E(b_1)$ in Equation 3 we get . . .

$$\bar{y}_\$ = \mu_\$ + r_{xy}SD_y\bar{z}_s$$

Equation 6

. . . where $\bar{y}_\$$ is the average dollar value of the work accomplished by those selected. Of course, nothing is free, including personnel selection tests. Subtracting out the cost of testing (C) an applicant we get an even better estimate of total dollar value added per applicant selected of . . .

$$\bar{y}_\$ = \mu_\$ + r_{xy} SD_y \bar{z}_s - C$$

Equation 7

Making a final change to reflect the number of applicants selected (N_s) and tested (N_a) we get the total dollar value added from N_s newcomers selected from N_a applicants:

$$N_s \bar{y}_\$ = N_s(\mu_\$ + r_{xy} SD_y \bar{z}_s) - N_a C$$
$$or$$
$$U_{total} = N_s(\mu_\$ + r_{xy} SD_y \bar{z}_s) - N_a C$$

Equation 8

Note, Equation 3 toEquation 7 focus on the total dollar value added from work performance of those selected using some personnel selection system. They do not tell us how much of that performance was due to use of the personnel selection system. The portion of the total dollar value added by those selected due to the personnel selection system is usually called the utility of that selection system. The utility or dollar value added to the firm due to use of the personnel selection system by the N_s individuals selected can be estimated by subtracting $\mu_\$$ from both sides of Equation 8. Recall $\mu_\$$ is the dollar value of work performance the firm expected to get when it chose applicants at random (i.e., what it would have received without use of the selection test), hence, $\bar{y}_\$ - \mu_\$$ is equal to the dollar performance gain resulting from use of the selection procedure, or . . .

$$N_s(\bar{y}_\$ - \mu_\$) = N_s r_{xy} SD_y \bar{z}_s - N_a C$$

Equation 9

Equation 9 is often written as . . .

$$\Delta U = N_s r_{xy} SD_y \bar{z}_s - N_a C$$

Equation 10

. . . where ΔU is the change in utility in dollar terms expected due to use of the personnel selection system to select N_s new hires from N_a applicants (see Boudreau, 1991, for BCG model extensions that reflect average job tenure, depreciation, marginal tax rates, etc.).

In sum, Equation 10 tells us the net dollar impact a selection system has, while Equation 8 equals the gross or total expected dollar impact of selecting N_s new hires from N_a applicants.[5] Cronbach and Glaser (1965) extended Brogden's (1949) model to two-stage and multi-stage selection, fixed treatment selection, placement, and classification decision situations (as one might imagine, the formulae get more complicated). Regardless, Equation 8 and Equation 10 show that it is r_{xy} (or $R_{y-x_1\,x_2 \ldots x_k}$ when multiple predictors are used) that is linearly related to actual dollar impact on the firm, not r_{xy}^2 or $R_{y-x_1x_2}^2$. This provides the basis for saying that a model characterized by $r_{xy} = .30$ explains 30% of the economic utility available to be predicted in the criterion Y, even though it explains only 9% of variance in Y. When $r_{xy} = .30$, each increase of 1 SD in standardized test score is expected to be paired with .30(SD_y) increase in economic utility.

Some BCG Model Implications. A number of implications follow from the BCG model that are not immediately obvious from Equation 10. First, it is not immediately obvious that a model yielding $\Delta R^2 = R_{y-x_1x_2}^2 - r_{y-x_1}^2 = .35 - .10 = .25$ will yield higher utility than a model that yields $\Delta R^2 = R_{y-x_1x_2}^2 - r_{y-x_1}^2 = .75 - .50 = .25$, even though both incrementally increased "variance explained in $y_\$$ by 25%." This is true because:

[5] See Russell, Colella, & Bobko (1993) for an in depth discussion of the different implications of Equations 8 & 10.

i. when $\Delta R^2 = R^2_{y-x_1x_2} - r^2_{y-x_1} = .35 - .10 = .25$, ΔR = $\sqrt{.35} - \sqrt{.10} = .59 - .32 \sim .27$; while when . . .

ii. $\Delta R^2 = R^2_{y-x_1x_2} - r^2_{y-x_1} = .75 - .50 = .25$, ΔR = $\sqrt{.75} - \sqrt{.50} = .87 - .71 \sim .16$.

In other words, two incremental advances in strategic management theory which both increase variance explained in firm profit when $Y_\$$ by 25% will not result in equal increases in actual economic value to the firm. **Error! Reference source not found.** below plots how ΔR changes as the base model r^2_{xy} increases from 0.00 for ΔR^2 = .25, .16, .09, and .04. In addition to Brogden's (1949) derivation, **Error! Reference source not found.** clearly shows R^2's deficiency as an index of model usefulness or prediction strength in organizational settings. "Incremental increase in percentage of $Y_\$$ variance explained" (r^2_{xy} or $R^2_{y-x_1x_2\ldots\, x_k}$) has a nonlinear relationship with organizational outcomes predicted by strategic theory <u>and</u> that nonlinear relationship changes in a nonlinear way as $R^2_{y-x_1x_2\ldots\, x_k}$ for the base model increases. To draw a specific contrast, consider that $r^2_{xy} = .09$ means $r_{xy} = .30$, or 30% of one standard deviation of dollar value ($SD_{Y\$}$) is gained for every 1 SD increase in X. However, if $\Delta R^2 = .09$ when X is added to a group of pre-existing predictors whose "base" model yielded R^2_{base} = .49, then $\Delta R = \sqrt{.58} - \sqrt{.49} = .76 - .70 = .06$, and we can only say X incrementally increased prediction utility by 6%. ΔR^2 will generally be smaller than ΔR when $R^2_{bas\,e}$ is small, though as Figure 1 shows, as R^2_{base} increases, ΔR^2 rapidly becomes larger than ΔR.

Insert Figure 1 about here

Second, Equation 8 andEquation 10 describe the expected total and incremental dollar value added by personnel selected using some personnel selection system. Most HRM decision makers will not be interested in forecasted expected dollar performance $\hat{Y}_{\$_i}$ for some individual applicant "i." Each "application" of the system is to some number of job applicants (N_a), and the value added to the firm comes from the performance realized from the entire subset of applicants selected by the selection system (N_s). Parallel application of the BCG model at strategic levels would occur in large corporations containing multiple strategic business units (SBUs), where a central authority would impose common strategic interventions on SBUs. Just as the value-added of a personnel selection system is realized from the performance of each individual selected using that personnel selection, the value-added of a strategic intervention would result from the incremental increase in performance of each individual SBU in which the strategic intervention was applied.

However, in contrast to most HRM applications, strategic decision makers will also be interested in point estimates of $\hat{Y}_{\$_i}$, or the forecasted dollar outcome of some strategic intervention X (or array of strategic interventions $X_1 \ldots X_k$). In other words, while HRM professionals will not be particularly interested in a point estimate of the expected performance for any individual applicant, CEOs and other strategic decision makers will be very interested in both dollar impact point estimates ($\hat{Y}_{\$_i}$) and prediction intervals around those point estimates for strategic interventions in their firms. L. Kevin Cox, V.P. of Human Resources at American Express, will have little interest in the forecasted performance of any one newly hired call center employee ($\hat{Y}_{\$_i}$ obtained after applicant i's standardized test score Z_i is plugged into Equation 7). In contrast, Kenneth I. Chenault, American Express' chairman and CEO, will have great interest in both the point estimate profit forecast and its associated prediction interval for the various alternate strategic interventions he might be considering.

Finally, before applying the BCG model to strategic management research results reported in a recent issue of the Academy of Management Journal, it should be noted that the BCG model is not limited to derivations using OLS optimization methods. Many strategic research circumstances and accompanying designs violate one or more OLS assumptions (e.g., normality of error terms). When these assumptions are violated in known ways, alternate optimization procedures are applied and hypotheses tested using different probability density functions (e.g., Logit, Probit, feasible generalized least squares, etc.). Regardless, all procedures I have encountered ultimately yield one or more models of the kind described in Equation 1 & Equation 2. Regardless of the optimization procedure used to estimate equation parameters, the resulting model can estimate $\hat{Y}_{\$_i}$ expected from any given strategic intervention – the BCG model is not OLS dependent.

An Example from George (2005).

I will now demonstrate how the BCG model might be applied using results reported by George (2005) in a recent issue of the Academy of Management Journal. Note, I picked this article entirely at random and as best I can tell, the author used appropriate methods and drew appropriate inferences. I chose this article only to illustrate how the BCG model might be applied to help strategic decision makers in privately held firms (the population George addressed) estimate expected dollar returns if they decide to use George's results to increase their profit. How close actual dollar returns are to expected dollar returns will constitute the acid test of how "good" George's (2005) model is if one adheres to Lewin's inverse, i.e., one believes models are not very good unless they are useful.

George (2005) examined relationships between profit and sets of behavioral and resource constraint measures in a sample of 900 privately held firms. George's base model predicted profit from lagged measures of firm size, industry profitability, number of competitors, competitor size, industry complexity, number of plants, firm age, and whether the firm was family managed or not. He used a cross sectional feasible generalized least squares procedures to estimate Equation 1 coefficients and to control for heteroskedasticity and autocorrelations, which yield log-likelihood estimates of prediction accuracy.[6] George did report $R^2_{y-x_1\,x_2\ldots x_k}$ his Models 1 (base), 2 (main effect), and 7 (full model including interaction effects) using a time series fixed effect analysis. The respective $R^2_{y-x_1\,x_2\ldots x_k}$ were .30, .41, and .56 for Models 1, 2, and 7, respectively. Again, using traditional interpretations of coefficients of determination, Model 1 explains 30% of the variance in profit, while the main effects model yields an 11% increase and the full interactive model adds a 26% increase in variance explained. However, $R_{y-x_1\,x_2\ldots x_8} = \sqrt{.30} =$ .55 for the base model, $R_{y-x_1\,x_2\ldots x_{16}} = \sqrt{.41} =$ .64 for the main effect model, and $R_{y-x_1\,x_2\ldots x_{32}} = \sqrt{.56} =$ .75 for the full interactive model. Application of BCG model logic indicates the main effect model (Model 2) increased expected dollar utility by 16% ($\frac{.64}{.55} = 1.16$) relative to the base model, and the full interactive model (Model 7) increased expected dollar utility by 36% ($\frac{.75}{.55}$ = 1.36).

While ΔR and ΔR^2 may seem fairly close in these instances, recall that ΔR accurately reflects the expected dollar impact of strategic interventions suggested by George's (2005) models. In industries where profit margins are in the low single digits (e.g., for profit health care

[6] Had George reported the log-likelihood of the null model (LL_{null}), I could have estimated the Cox and Snell approximation of R generated by OLS regression for each of George's Models 1-7, where $R_{logistic} = \sqrt{\frac{-2LL_{null} - 2LL_k}{-2LL_{null}}}$, LL_{null} is the log likelihood of a model containing just a constant (i.e., a function of the average profit across the entire span of the study), and LL_k is the log likelihood of the model containing k predictors.

margins typically range from 1-4%), it is important to know exactly what incremental profit improvement is expected from planned strategic changes. $R_{y-x_1\,x_2\ldots x_{32}} = .75$ reflects the fact that George's full interactive Model 7 accounts for 75% of the dollar profit available to be explained within his sample of 900 firms. Of course, any point estimate forecast of dollar profit expected IF one of George's privately held firms were to act on his results would also have to subtract any costs associated with making the desired strategic changes (i.e., the N_aC value in Equation 8). Further, only 19 of 32 predictors significantly contributed to Model 7, and any actual implementation of Model 7 would be characterized by whatever lower $R_{y-x_1\,x_2\ldots x_{19}}$ is associated with that reduced 19 predictor model. Of course, coefficients estimated for these 19 predictors would likely differ substantially from those George (2005) reported for Model 7 due to change in effects of multicolinearity between the 32 and 19 predictor models (e.g., the coefficient for industry profitability ranged from -20.74 for Model 7 to -6.49 for main effect Model 2).

Concern II: Causal Language and Nudity among Strategy Scholars

My second concern stems from the extensive use of "causal language" in the absence of experimental or quasi-experimental designs in the strategic management literature. George (2005), like most other strategic management scholars, liberally used causal language throughout his theory development and interpretation of results. Note, I am not immune from this criticism either. I had a heated discussion with two coauthors on this issue before bowing to their pressure in titling an article "The effect of cultural distance on . . . " (Tihanyi et a., 2004),when I knew full well that no evidence of cultural distance causing anything was present in the analyses we reported.

Doctoral students routinely brought this up during methods seminars over the last 20 years, asking "Why, if correlation does not mean causation, can {famous strategy scholar} say 'these results strongly suggest X influences Y, supporting the scholarly theory and hypotheses I laid out in the introduction'?" My only answer to them is that, apparently, this is one of those examples of the Emperor not knowing he was naked. If strong causal inferences were justified by George's (2005) results, more than 900 CEOs of privately held firms would have been vying for seats at his 2003 Academy of Management presentation of these results.

Unfortunately, wishing does not make it so. Just because cross sectional, correlational results are "consistent" with a causal model does not mean strong inferences of causality are justified, e.g., inferences strong enough to justify changing a firm's strategy. So, what can be done?

Cross-Validation. First, strategy scholars could routinely cross-validate their results. Specifically, it is highly unlikely that George's (2005) Model 7 actually explains 75% of the dollar utility available to be explained in privately held firm profits. We are all familiar with the robust beauty and power of statistically optimized prediction equations.[7] Cross-validation is one way to account for the fact that actual predictive power will be attenuated by BOTH sampling error in Geroge's (2005) N = 900 sample and sampling error in whatever collection of SBU's one might apply George's findings to. Efron and Tibshirani (1997) proved that the .632 bootstrap method of cross-validation is the most efficient means of estimating cross-validities, while Dean and Russell (2001) demonstrated how it could be applied in management research.

---

[7] Dawes and Corrigan (1974) demonstrated in a Monte Carlo simulation that when X→Y relationships are monotonic, simple additive models chosen at random predict on average 92% of the variance in Y that would have been explained IF one had used the actual nonlinear model that originally generated the data. Given the paucity of non-monotonic relationships (i.e., U or inverted-U shaped relationships) in management research, this is yet another source of R inflation for incorrect models.

Applied to George's (2005) analyses, the .632 bootstrap cross-validity estimation procedure would have (for example) generated 1000 samples of 900 firms with replacement from George's original sample of 900, estimated each of the models in each of the 1000 bootstrap samples, then cross-validated each model on the approximately 331 (331.2 = 900 - .632{900}) firms that had not been included in each bootstrap sample. George's (2005) estimates of model coefficients reported in his Table 2 would have still been "best" estimates, though the FGLS log-likelihoods and time series fixed effects $R^2_{y-x_1\,x_2\ldots x_{32}}$ (reported in the text) would have been average log-likelihoods and $\bar{R}^2_{y-x_1\,x_2\ldots x_{32}}$ obtained when the models were applied to the 1000 "hold out" samples of ~ N_k = 331. These average log-likelihoods and $\bar{R}^2_{y-x_1\,x_2\ldots x_{32}}$ constitute the best estimate of how the log-likelihoods and $R^2_{y-x_1\,x_2\ldots x_{32}}$ Georges (2005) reported will be attenuated when used to make forecasts in future samples.

Cross-validation (regardless of method) should rein in reporting of effect sizes that are inflated due to capitalization on chance sampling error. Unfortunately, it will not solve the causality problem by itself.

Quasi-experimental Consulting & Case Studies. Criticizing use of causal language in the absence of experimental or quasi-experimental designs in macro-management research arenas is easy, while solutions remain elusive. I am very aware of how easy it is to use random effects and fixed effect experimental and quasi-experimental designs in the more micro-oriented management research arenas. With advent of the internet, I have routinely had access to large heterogeneous and homogenous samples in my personnel selection research and, increasingly, have opportunities to collaborate with firms implementing quasi-experimental designs to assess alternate HRM interventions. And yet, at some point in each of these projects I always recall my OT doctoral seminar professor over 30 years ago describing reverently how Joan Woodward's groundbreaking organizational research of the 1950's shattered everyone's notion of what was possible by obtaining a sample of 58 companies (Woodward, 1965). If I dwell on it too long, I recall Karl Weick's chapter titled "Laboratory experiments with organizations" (Weick, 1965) and revisit all over again my ~ 1978 career choice to "go micro" because macro-organizational research was just too hard to do!

Yet, in just the last 10 years I have personally experienced another way to validate and test causal inferences. Working first as a member of ePredix Inc.'s and subsequently (post merger) PreVisor Inc.'s Technical Advisory Board has permitted routine access to samples ranging from 5,000 to a high of > 87,000 in predictive validity designs. Select client organizations have permitted use of quasi-experimental designs to assess the effects of different alignments of HR systems, job requirements, and labor market conditions. Assessing the effects of on-line versus traditional proctored paper and pencil completion of personnel selection tests was accomplished in this manner. A similar approach would yield stronger assessments of causal paths hypothesized by George (2005).

Specifically, George's (2005) full interactive Model 7 coefficients indicate one unit changes in. . .

1. industry profitability;
2. the product of high discretion slack and complexity;
3. the logarithm of firm sales;
4. the product of resource demand and complexity;
5. resource demand; and,
6. number of plants

. . . will all effect profit by more than $1M (some negatively, some positively). A very simple test of George's (2005) causal assertions would be to compare forecasted profit change to actual profit change in firms that actually made strategic changes consistent with George's (2005) Model 7. Unfortunately, the strongest contributor to Model 7, Industry Profitability, is not something typically within the control of strategic decision makers, though Model 7 suggests entrepreneurs making initial point of entry decisions would do well to chose industries populated by publicly held corporations with low ROA - every one point decrement in average publicly held corporation ROA is accompanied by an expected increase of $20.74M in profit by the privately held firm.

Other predictors can be influenced by strategic decision making. For example, the average log of firm sales (within and across 900 privately held firms from 1994-97) was 16.54 (SD = 2.21), so average firm sales were ~ $15.25M. The 2.39 coefficient reported for Firm Size in Model 7 means privately held firms in this sample increased profit by $2.39M when they went from $15.25M in sales (ln{$15.25M} = 16.54) to $41.45M in sales (ln{$41.45M} = 17.54). Firm which increased sales to $112.67M gained another $2.39M in profit (ln{$112.67M} = 18.54). Jumping to $306.28M in sales yielded yet another $2.39M in profit ln{$306.28M} = 19.54). These results suggest growing larger is not the easy way to increase profits in this sample of privately held firms – strategic decision makers should probably look elsewhere for ways to enhance profit.

Next, Model 7 indicated a one unit increase in the interaction of High-Discretion Slack and Complexity is expected to yield a $2.5M profit increase, and gets at the core of George's (2005) theoretical contribution. Industry Complexity was operationalized as the sum of squared market shares of publicly traded firms in the 4-digit SEC sector and, again, is not likely to be easily effected by strategic decision makers in privately held firms. In contrast, High-Discretion Slack was operationalized as the level of cash reserves in a given year. Cash reserves could be influenced in a number of ways by strategic decision makers, e.g., through decisions to retain earnings instead of paying dividends. George's (2005) Model 7 suggests every increase of $1M in annual cash reserves multiplied by the sum of squared outstanding shares issued by publicly held 4-digit SEC code peers yields a $2.5M increase in profit. In other words, if the sum of squared outstanding peer competitors' shares issues was 1,000, a $1,000 increase in annual cash reserves is expected to increase annual profit by $1M.

In contrast, Model 7 predicted Resource Demand ($X_{\text{Resource Demand}}$ = five days sales plus accounts receivable plus inventory minus accounts payable) is a multi-edged sword – increasing Resource Demand by $1M is expected to change profit as follows:

Main effect:	+$1.5M
Squared main Effect ($X^2_{resource\ \ demand}$):	-$.0002M($X^2_{resource\ \ demand}$)
Complexity Interaction:	-$1.58M($\sum_{i=1}^{k} Z^2_{shares}$)
Age Interaction with $X^2_{resource\ \ demand}$:	+$.000001M($X^2_{resource\ \ demand}$)($X_{firm\ \ age}$)

Due to the squared terms ($X^2_{resource\ \ demand}$), total expected effect on profit will decrease exponentially as Resource Demand increases - the strategic choice to raise Resource Demand by $2M would have a net effect on Profit of 2[$1.5M -$1.58M($\sum_{i=1}^{k} Z^2_{shares}$)] – 2^2[$.000001M($X_{firm\ \ age}$) + $.0002M], while raising it by $4M would change profit by 4[$1.5M -$1.58M($\sum_{i=1}^{k} Z^2_{shares}$)] – 4^2[$.000001M($X_{firm\ \ age}$) + $.0002M]. As the product $1.58M($\sum_{i=1}^{k} Z^2_{shares}$) will always be larger than $1.5M and the remaining portion of the effect gets exponentially more negative as Resource Demand increases, lower resource demand should always yield greater profit.

Given these observations drawn from Geroge's (2005) Model 7, strong tests of causal inferences about High-Discretion Slack and Resource Demand will occur when one compares actual change in profit to expected change in profit after effecting change in privately held firms' High-Discretion Slack and Resource Demand. This involves something more than access to the Dun & Bradstreet database matched to data from <u>Ward's Business Directory of Privately Held Firms</u>. One would start by identifying which of George's (2005) 900 privately held firms is expected to benefit most from changes in controllable strategic decisions (e.g., Resource Demand, High-Discretion Slack, Low-Discretion Slack, and Resource Availability). Soliciting participation by these firms in a field study would, at a minimum, permit the monitoring of any change in profit paired with any changes in Resource Demand, High-Discretion Slack, Low-Discretion Slack, or Resource Availability that might happen to occur. If initial results are promising and initial forecasted changes in profit are realized, it might even open doors to the possibility of active strategic interventions, i.e., comparison of actual profit to forecasted profit when strategic changes were made <u>based on</u> George's (2005) promising initial scholarship with privately held firms .

<u>Conclusion</u>

In sum, my major concerns from reading the strategic management literature were two-fold. First, strategy scholars seem overly enamored with coefficients of determination (r^2_{xy} or $R^2_{y-x_1x_2}$), a statistic that is not linearly related to the "usefulness" of the underlying theory or model. A simple reporting and interpretive change focusing on r_{xy} and $R_{y-x_1x_2\ldots x_k}$ will resolve this problem. Additional focus on generating models that might actually be used by strategic decision makers (e.g., the slimmed down 19 predictor version of George's, 2005, Model 7) and cross-validation will yield improved estimates of a model's actual usefulness and, for inverted Lewinians like myself, quality. Second, strong causal language is not justified in virtually every piece of strategy scholarship I read. More proactive involvement in real organizational settings, or “quasi-experimental field studies,” is needed before strong causal inferences are justified. More applied field research is needed to move strategy beyond its current state of development.

References

Boudreau, J.W. (1991). Utility analysis for decisions in human resource management. In Dunnette, M.D. & Hough, LM. (eds.), Handbook of Industrial and Organizational Psychology 2nd ed. (Vol. 2, pp. 621-745). Palo Alto, CA: Consulting Psychologists Press.

Brogden, H.E. (1949). When testing pays off. Personnel Psychology, 2, 171-185.

Cronbach, L.J. & Gleser, G.C. (1965). Psychological Tests and Personnel Decisions (2nd Ed.). Urbana: University of Illinois Press.

Dawes, R.M. & Corrigan, B. (1974). Linear models in decision making. Psychological Bulletin, 85, 90-106.

Dean, M.A. & Russell, C.J. (2001, August). Bootstrap cross-validation efficiencies in personnel selection. Presented at the annual Academy of Management meetings, Washington, D.C.

Efron, B. & Tibshirani, R. (1997). Improvements on cross-validation: The .632+ bootstrap method. Journal of the American Statistical Association, 92, 548-560.

George, G. (2005). Slack resources and the performance of privately held firms. Academy of Management Journal, 41, 661-676.

Gulati, R. (2007). Tent poles, tribalism, and boundary spanning: The rigor-relevance debate in management research. Academy of Management Journal, 50, 775-782.

Kaplan, R. & Norton, D. (2005). The office of strategy management. Harvard Business Review, October, 82-80.

Ketchen, D.J.Jr., Combs, J.G., Russell, C.J., Shook, C., Dean, M.A., Runge, J., Lohrke, F.T., Naumann, S.E., Haptonstahl, D.E., Baker, R., Beckstein, B.A., Handler, C., Honig, H., &

Lamoreaux, S. (1997). Organizational configurations and performance: A meta-analysis. Academy of Management Journal, 40, 223-240.

Lewin, K. (1945). The Research Center for Group Dynamics at the Massachusetts Institute of Technology, Sociometry, 8, 126-136.

Russell, C.J. (1985). Individual decision processes in an assessment center. Journal of Applied Psychology, 70, 737-746.

Russell, C.J. (2001). A longitudinal study of top-level executive performance. Journal of Applied Psychology, 6, 510-517.

Russell, C.J., Colella, A., & Bobko, P. (1993). Expanding the context of utility: The strategic impact of personnel selection. Personnel Psychology, 46, 781-801.

Russell, C.J. & Dean, M.A. (2000). To log or not to log: Bootstrap as an alternative to parametric estimation of moderation effects in the presence of skewed dependent variables. Organizational Research Methods, 3, 167-185.

Russell, C.J., Settoon, R.P., McGrath, R., Blanton, A.E., Kidwell, R.E., Lohrke, F.T., Scifries, E.L., & Danforth, G.W. (1994). Investigator characteristics as moderators of selection research: A meta-analysis. Journal of Applied Psychology, 79, 163-170.

Tihanyi, L., Griffith, D. A., and Russell, C. J. (2004). The effect of cultural distance on entry mode choice, international diversification, and MNE performance: A meta-analysis. Journal of International Business Studies, 36, 270-283.

Weick, K.E. (1965). Laboratory experimentation with organizations. In March, J. (ed.), Handbook of Organizations (pp. 194-260). Chicago, IL: Rand McNally.

Woodward, J. (1965). Industrial Organization: Theory and Practice. New York: Oxford University Press.

Figure 1: Changing ΔR When ΔR^2 = .25, .16, .09, and .04 Across All Baseline Models