Check for updates

Organizational Research Methods
14(2) 287-305
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1094428110369842
http://orm.sagepub.com
SAGE

Methodological Urban Legends: The Misuse of Statistical Control Variables

Paul E. Spector[1] and Michael T. Brannick[1]

Abstract

The automatic or blind inclusion of control variables in multiple regression and other analyses, intended to purify observed relationships among variables of interest, is widespread and can be considered an example of practice based on a methodological urban legend. Inclusion of such variables in most cases implicitly assumes that the control variables are somehow either contaminating the measurement of the variables of interest or affecting the underlying constructs, thus distorting observed relationships among them. There are, however, a number of alternative mechanisms that would produce the same statistical results, thus throwing into question whether inclusion of control variables has led to more or less accurate interpretation of results. The authors propose that researchers should be explicit rather than implicit regarding the role of control variables and match hypotheses precisely to both the choice of variables and the choice of analyses. The authors further propose that researchers avoid testing models in which demographic variables serve as proxies for variables that are of real theoretical interest in their data.

Keywords

field research methods, multiple regression, survey research, statistical control

The use of statistical control variables in nonexperimental research is routine and widespread. Atinc and Simmering (2008) found in their review of the literature that papers had a mean of 7.7 and 3.7 control variables in macro-organizational and micro-organizational studies, respectively. It has become standard practice in questionnaire studies to test hypotheses relating theoretical antecedents and consequences (independent and dependent variables) using hierarchical multiple regression with control variables entered in the first step. Often the zero-order correlations between the main variables of interest in such studies are ignored, and may go unmentioned when conclusions are drawn, despite the fact that hypotheses were bivariate and not conditional on the controls. Frequently ignored too are discrepancies between zero-order correlations and regression results that suggest

[1] Department of Psychology, University of South Florida, Tampa, USA

Corresponding Author:
Paul E. Spector, Department of Psychology, University of South Florida, PCD 4118, Tampa, FL 33620, USA
Email: pspector@usf.edu

suppression or other complex interplays among variables. Such practices can lead to confusion in the interpretation of results, as was well documented by Breaugh (2008).

The distinguishing feature of control variables is that they are considered extraneous variables that are not linked to the hypotheses and theories being tested. Their role is assumed to be confounding, that is, producing distortions in observed relationships. Researchers clearly delineate some variables as being merely controls or variables of no particular theoretical interest that need to be somehow removed in their effects on the study. Rather than being included on the basis of theory, control variables are often entered with limited (or even no) comment, as if the controls have somehow, almost magically, purified the results, revealing the true relationships among underlying constructs of interest that were distorted by the action of the control variables. This is assumed with often little concern about the existence and nature of mechanisms linking control variables and the variables of interest. Unfortunately, the nature of such mechanisms is critical to determining what inclusion of controls actually does to an analysis and to conclusions based on that analysis. For this reason, one can find admonitions to be suspicious of results with such blindly applied statistical controls by respected methodologists such as Meehl (1970, 1971) and Pedhazur (1997, pp. 170-172). Such cautions are far from new and can be found as far back as the 1920s (Burks, 1926) suggesting that intemperate use of statistical controls has been common practice for generations.

The implicit belief that statistical controls can yield more accurate estimates of relationships among variables of interest, which we will call the "purification principle," is so widespread, and is so accepted in practice, that we argue it qualifies as a methodological urban legend—something accepted without question because researchers and reviewers of their work have seen it used so often that they do not question the validity of the approach. It is a legend because of two things. First, it is based on a kernel of truth, that is, that statistical controls are able to adjust relationships between variables for the action of other variables. Second, although there are times a control variable can appropriately adjust for the effects of an extraneous variable, it is unable to accomplish this goal in all circumstances, and the nature of what it can actually test is quite limited. More specifically, statistical control is based on certain implicit assumptions about the underlying role of control variables on either the observed measures or the underlying constructs of interest. If that role is incorrect, then conclusions based on analyses using those controls will likely be incorrect as well. It is an urban legend that control variables will always have a role that renders their inclusion appropriate and that the purification principle is generally correct.

Our purpose in this article is threefold. First, we will briefly review the nature of the problem with the purification principle and why it can lead to erroneous inference. We will include discussion of the implicit assumptions one makes about the role of control variables when including them in an analysis and what analyses with controls can actually test. Our main focus will be on nonexperimental research using multiple regression, although the problems we note will apply to other statistical control approaches, such as analysis of covariance and structural equation modeling (SEM). We differ from others who have discussed this issue (e.g., Breaugh, 2008; Meehl, 1971) in focusing on the distinction between factors that affect the measurement of a construct versus the underling construct itself. Second, we will suggest a more focused and theory-based use of controls. Rather than suggesting that controls are to be avoided altogether, we suggest that they have an important use in testing competing hypotheses involving controls. We go beyond other critics of improper use of controls (e.g., Becker, 2005; Breaugh, 2008) in focusing on the use of controls to test competing hypotheses. Third, we discuss the appropriate use of demographic variables and how they are inappropriately used as proxies for underlying variables that should be directly investigated. Control variables serve an important role in programmatic efforts to explain the reasons for our observed results. Our problem is not that controls are used but rather how they are used.

The Problem With the Purification Principle

Alternate Mechanisms Involving Control Variables

Statistical control variables are often used in an attempt to yield more accurate (purified) estimates of relationships among underlying theoretical constructs of interest. They are used routinely in tests of bivariate hypotheses as well as more complex multivariate hypotheses such as mediation or moderation. However, controls are typically used with insufficient discussion of why they were included and how their inclusion would lead to more accurate conclusions. Even when the role of control variables is noted, typically there is little or any evidence in support of such a role. Becker (2005) examined a sample of papers in which control variables were used, finding that in more than half of the cases, no explanation was provided for including one or more control variables, and in more than two thirds of cases, there was no evidence provided to justify inclusion of one or more control variables. Similarly, Atinc and Simmering (2008) found that the majority of papers in their review of the literature did not provide citations to support the inclusion of one or more control variables. These findings clearly show that in most cases, authors fail to make a compelling case for inclusion of control variables in their analyses, and in many cases they may be relying on the purification principle. Of course, it is also possible that researchers have used an exploratory or shotgun approach and "discovered" that variables of interest are only significantly related when "control" variables are also included in an analysis. Certainly, readers and reviewers should be on the lookout for such results that seem to be the product of suppression effects that are unexplained and perhaps unnoticed.

We looked closely at the justifications researchers typically give for including controls by reviewing all empirical papers from the two most recent issues of *Academy of Management Journal*, chosen because it is a top tier journal that publishes empirical papers in both the micro and macro areas. We do not target any of these authors for particular criticism, as they handled controls in a manner consistent with common practice and demands of the review process, and serve as exemplars of such practice. Of the 20 empirical articles published in October and December 2009, 18 described using control variables in the analysis of their nonexperimental data. Sixteen devoted a separate subsection of the method that listed the control variables and explained how they were operationalized. Sixteen of the articles included statements to justify why the controls were included. In the majority of cases, the language used suggested some sort of causal connection. Five papers noted that the control variable "affected" other variables in the study (Bidwell & Briscoe, 2009; Cuervo-Cazurra & Dau, 2009; Kennedy & Fiss, 2009; McDermott, Corredoira, & Kruse, 2009; Pil & Leana, 2009), for example, "we controlled for other factors that might affect firm profitability" (Cuervo-Cazurra & Dau, 2009, p. 1356). Four noted that the control variables might "influence" other variables (Hoobler, Wayne, & Lemmon, 2009; Lee, 2009; Marquis & Huang, 2009; Tzabbar, 2009), for example "I controlled for a host of variables that might influence policy dynamics" (Lee, 2009, p. 1256). Two used the term "impact" (Leana, Appelbaum, & Shevchuk, 2009; Madsen, 2009). Others used terminology that was perhaps a bit more vague about the role of controls variables, such as "Firm size plays an important role in explaining market returns" (He & Wang, 2009, p. 926), "We controlled for several variables that might account for relationships between our independent and dependent variables" (Shaw, Dineen, Fang, & Vellella, 2009).

Not only do most of these papers include statements suggesting a causal connection between control and other variables, but they express uncertainty about such a role, with eight of them qualifying their statements with a "may" or "might," and five others using words such as "expected," "possibility," "potential," and "thought to have." In almost all cases, there was little or no evidence provided that the control variables played the role suggested. A notable exception was Kennedy and Fiss (2009) who cited a source suggesting that their control variable (organization size) affected their dependent variable (speed of adoption).

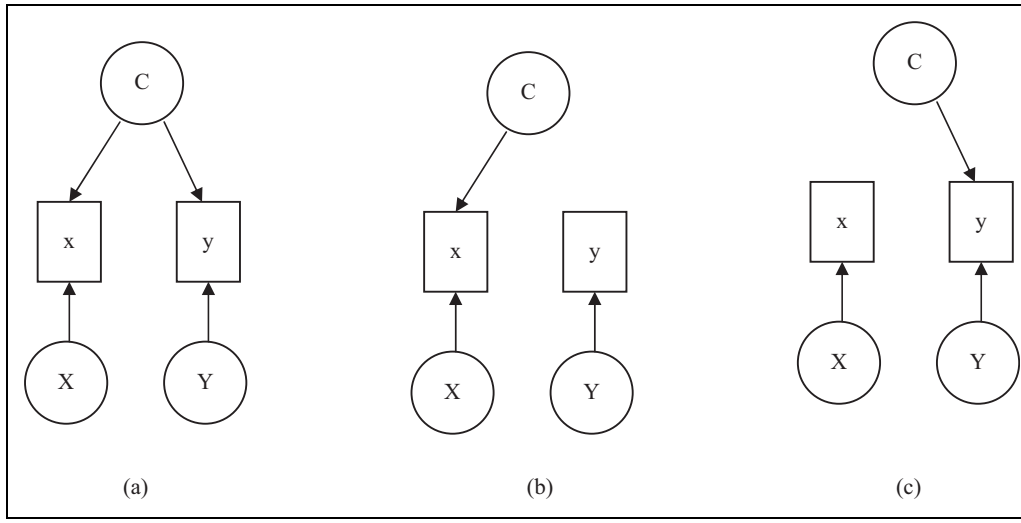

Figure 1. Models of contamination with a control variable affecting both measured variables (A), Just *X* (B), or Just *Y* (C)

As our analysis suggests researchers are concerned that the variable or variables being controlled are somehow influencing the variables of interest and are therefore producing a distorted view of the relationships among them. For example, a supervisor's liking for a person might inflate the supervisor's rating of that person's job performance across multiple dimensions. Correlations among those dimensions might well be influenced by liking, which in effect has contaminated ratings of performance. Thus, researchers might be tempted to control for liking when evaluating relationships among rating dimensions. Note, however, that whether it is reasonable to control liking in this instance depends on whether liking is in fact distorting observed relationships. If it is not (perhaps, liking is the *result* of good performance), treating liking as a control will lead to erroneous conclusions. This is because removing variance attributable to a control variable (liking) that is caused by a variable of interest (performance) will remove the effect you wish to study (relationships among performance dimensions) before testing the effect you wish to study, or "throwing out the baby with the bathwater," as noted by Spector, Zapf, Chen, and Frese (2000, p. 91).

In a general sense, researchers are interested in theoretical relationships among underlying constructs. The whole idea of hypothesis-driven research is that the researcher has used prior findings and theory to predict relationships among variables of interest, in some cases based on hypothetical explanatory mechanisms. Thus, researchers might study supposed antecedents or consequences of attitudes, behavior, cognitions, personality, and so forth. Given interest in the relationship between a proposed antecedent *X* and consequence *Y*, there can be concern that a third variable *C* has somehow affected the observed relationship. This can happen in a number of ways, on both the construct and measurement level, that is, *C* can affect the assessment but not the constructs (contamination), or it can affect the underlying constructs of *X* and *Y* themselves (spuriousness).

Contamination. Contamination occurs when a third (control) variable (*C*) influences the observed measures of interest (*X* and *Y*). *C* does not affect the underlying constructs but only the measures of them. This possibility is illustrated in Figure 1a. Theoretical constructs in this figure are represented by circles and observed variables are represented by squares. The arrows from the constructs to the observed variables represent factor loadings from the intended constructs to observed indicators of those constructs, *X* to *x* and *Y* to *y*. Contamination is indicated by the arrow from *C* to *x* and *y*. Contamination as we present it is much like what is considered method variance, that is, extraneous

features of the method that produce variance in observed variables (Campbell & Fiske, 1959). What distinguishes method variance from our broader concept of contamination is that the former is assumed to reside in the method, whereas we are not restricting the action of contamination to features of method. For example, suppose we wish to determine if the relationship between self-reported job satisfaction and job performance might be contaminated by the personality variable of neuroticism. Method variance could be manifested if all variables were assessed with the same self-report method. Contamination that is independent of possible method variance would be observed if neuroticism is assessed in a different way, for example, by having trained observers watch individuals and rate their personalities.

The effects of contamination can be seen on the components that comprise the correlation coefficient between two variables, covariance in the numerator and variances in the denominator. Following classical test theory, the variance in a sample of observed scores, say X, can be partitioned into true score $\mathrm{Var}(X_{\mathrm{T}})$ and error $\mathrm{Var}(X_{\mathrm{E}})$ components. If X is contaminated by variable C, this adds another variance component, as shown in equation 1 for X. Note that the variance of uncontaminated X would be equal to only the first two X_T and X_E components.

$$\mathrm{Var}(X_{\mathrm{Contaminated}}) = \mathrm{Var}(X_T) + \mathrm{Var}(X_E) + \mathrm{Var}(X_C) \tag{1}$$

Assuming that X_T, X_E, and X_C are independent, and given variances cannot be negative, $\mathrm{Var}(X_{\mathrm{Contaminated}})$ will be larger than $\mathrm{Var}(X)$ unless $\mathrm{Var}(X_C) = 0$. The same holds for Y. Therefore, in the correlation formula, the variance components in the denominator will be inflated due to C.

Contamination also affects covariance between X and Y. We can represent observed scores X and Y as linear combinations of T, E, and C, and then represent the covariance between X and Y as the sum of covariances among all possible pairs of terms across the two linear combinations (Nunnally & Bernstein, 1994). Since E is assumed to be independent of T and C, all covariance terms involving E are equal to 0. This leaves us with equation 2. As can be seen, if $C = 0$, the observed covariance is the covariance between X_T and Y_T. If there is contamination, then we add the remaining three terms. We assume that contamination is independent of T meaning the second and third terms to the right of the equal sign are equal to 0, so in the end, the observed covariance in the contaminated case is inflated by the last term in the equation that represents the covariance between the contamination components across X and Y.

$$\mathrm{Cov}(X_{\mathrm{Contaminated}} Y_{\mathrm{Contaminated}}) = \mathrm{Cov}(X_T Y_T) + \mathrm{Cov}(X_T Y_C) + \mathrm{Cov}(X_C Y_T) + \mathrm{Cov}(X_C Y_C). \tag{2}$$

If the contamination components are unrelated across the two variables (X and Y are contaminated by two different things), the correlation is attenuated as the contamination acts like additional error variance by inflating $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$. If X and Y are contaminated by the same thing, then the effects on the numerator will be larger than the effects on the denominator in the correlation formula, and the correlation is inflated.

Going back to our performance appraisal example, if ratings of performance dimensions (x and y in Figure 1a) are influenced by liking (C in Figure 1a), x and y are contaminated. Statistically controlling for liking will presumably remove the effect of that contamination from the relationship between rating dimensions and yield more accurate estimates of the relationships among underlying performance dimensions (X and Y in Figure 1a). It should be kept in mind, however, that whether it is reasonable to control for liking depends on whether contamination has or has not occurred. If contamination is not at play in the data, conclusions based on analyses using controls might be incorrect. For example, there is evidence that job performance can lead to supervisor liking of subordinates, in that high performers are better liked (Lowin & Craig, 1968; Varma, Denisi, & Peters, 1996). If supervisors like subordinates who perform well and dislike those who perform poorly, controlling liking will lead to an underestimate of the true correlations among performance dimensions. This

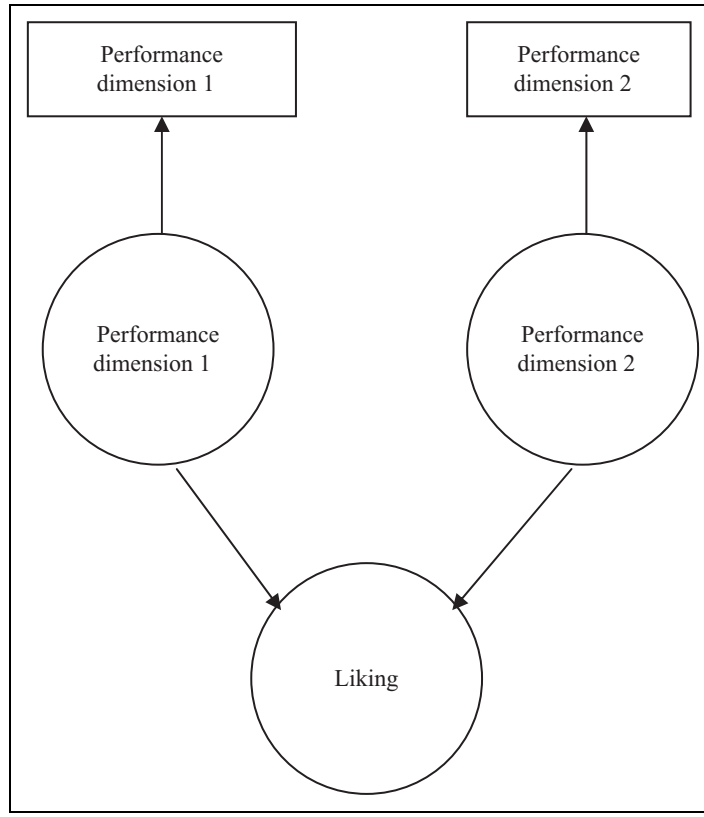

Figure 2. Model of performance leading to liking of subordinates by supervisors

possibility is illustrated in Figure 2. Liking is not a contaminant in this case, as it is not a cause of observations of performance. Rather, liking is the effect of performance, in that those performing well are liked because they perform well, and those performing poorly are disliked because they perform poorly. The observed correlation between performance dimensions correctly reflects their interrelationship. Statistical control of liking would result in inaccurate estimates of the relationships among performance dimensions, because it would remove part of the very effect of interest from estimates of the effect of interest.

Of course, a control variable can contaminate only X or only Y rather than both. This is illustrated in Figure 1b and c, showing that C can contaminate only one variable. If C contaminates X, for example, as shown in equation 2, the variance of X will be inflated by C. However, the covariance of X and Y will be unaffected because the right three terms in equation 2 will all be 0. Thus, the correlation will become smaller, as C acts like additional error variance in the equation—an independent factor that increases variance. Returning to our performance example, suppose we ask supervisors of sales representatives to indicate the representatives' sales volume (Y) and rate their cooperativeness (X). If those supervisors receive weekly objective sales data on each subordinate, their indication of sales volume might be based mainly on objective data and be unbiased by liking (C). The rating of cooperativeness is unlikely to be so immune, and so liking might contaminate the latter and not the former dimension. This would serve to render the cooperativeness rating as less valid than the sales rating because it is contaminated. Thus, the correlation between the two dimensions would be attenuated. Note if we had measures of X, Y, and C in this example, the use of regression would result in suppression. Variable X relates to Y, and C relates to X and not Y. Including C in

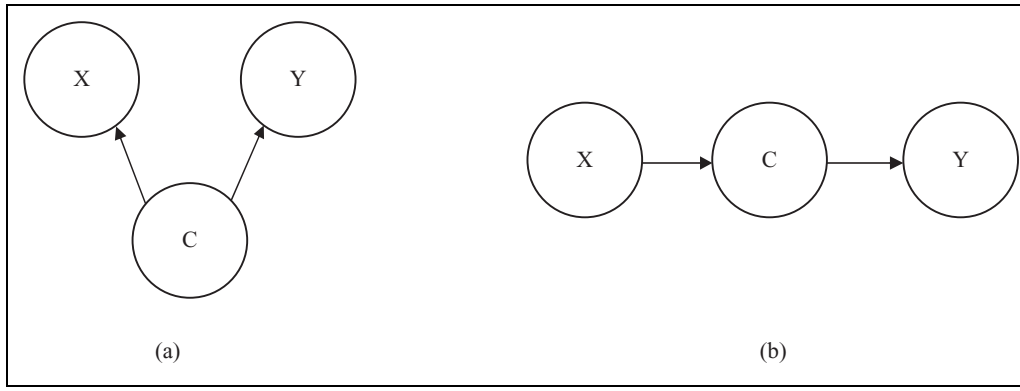

Figure 3. Models of spuriousness (A) and mediation (B)

a regression analysis will improve *X*'s ability to predict *Y*, even though *C* itself cannot predict *Y*. If *C* was related to *Y* and not *X*, however, both *X* and *C* would add independent incremental variance to the regression equation. The slope relating *Y* to *X* would not be affected by the addition of *C*, but the standard error of that slope would be reduced. Thus, the inclusion of *C* in a regression equation will have different effects depending on whether it contaminates *X* or *Y* (Figure 1b and c is asymmetric in their effects on the regression equation). This implication for purification does not appear to always occur to authors who include controls or reviewers who ask them to add controls based on the assumption that the controlling variable might contaminate only the dependent or outcome variable.

Spuriousness and other causal connections. There can be many reasons that observed variables relate to one another other than contamination of measurement. In other words, there can be a variety of connections among underlying constructs themselves. The inclusion of control variables that have such connections with variables of interest can be problematic as they can result in removal of the effects of interest from tests of the effects of interest. Thus, rather than purifying results, such practices render interpretation of such results incorrect.

One possible situation, illustrated in Figure 3a, is when the control variable *C* is a cause of both *X and Y*, which are not themselves causally related directly. This situation, in which two variables are related because of a common underlying cause, is spuriousness. If a potential control variable *C* is in reality a cause of variables *X* and *Y*, inclusion of *C* and *X* together in a regression analysis will provide an indication of the relationship between *X* and *Y*, removing the effect of *C* from *X*. If the relationship between *X* and *Y* is entirely due to *C* (i.e., is entirely spurious), the regression coefficient for *X* is expected to be 0 when *C* is also entered into the analysis. A more complete interpretation might compare the regression coefficients for *X* when it is entered alone versus accompanied by *C*. The extent of spuriousness would be revealed in the reduction in the standardized regression coefficient for *X*. If *X* and *Y* are not related in the first place, there is no (simple linear) relationship to explain.

Suppose, however, that the researcher is merely hypothesizing that the constructs are associated (i.e., there is a nonzero correlation between *X* and *Y* in the population). The most direct test of this hypothesis is the zero-order correlation or regression of *Y* on *X* alone. An analysis that includes a control variable *C* is testing a different hypothesis, for example, spuriousness as depicted in Figure 3a. If spuriousness is in fact the case, we will likely find that *X* is not significant once *C* is controlled and will conclude that *X* and *Y* are not related. This would not be a test of the hypothesis as stated, which only involves *X* and *Y*. Rather, it is the test of a more complex hypothesis that involves *C*. Unfortunately, in most cases in which controls are used, it is the relationship between *X* and *Y* alone that is hypothesized, and not the more complex hypothesis involving controls. Thus, in such

cases, the researcher's actual test (multiple regression analysis test of the significance of the regression weight for *X*) corresponds, not to the explicit hypothesis, but rather to some other hypothesis that is never clearly stated.

As we already illustrated, one way in which the control variable may influence the results is through contamination rather than spuriousness, and contamination will have different influences on the results depending on whether *C* influences, *X*, *Y*, or both. The second way the hypothesis is ill-defined is that there are different causal connections among *X*, *Y*, and *C*, which will produce the same pattern of observed results. As pointed out by MacKinnon, Krull, and Lockwood (2000), the "test" for spuriousness is identical to the "test" for mediation (MacKinnon et al., 2000, used the term confounding rather than spuriousness), and there is no statistical test for deciding which is correct. Two of several equivalent (in the sense of producing fit to the data) alternative models are illustrated in Figure 3a (spuriousness) and 3b (mediation). For both illustrated models, we expect *X* to be nonsignificant when both *X* and *C* are entered in the same multiple regression model as predictors of *Y*. The point is that the regression analysis cannot make the researcher's theoretical position (the hypothesis of interest) explicit, and in fact the regression results are not helpful in distinguishing among a number of different theoretical possibilities. This is one reason that we cannot draw confident causal conclusions from cross-sectional survey data.

The necessity for being clear about assumptions concerning controls. The accuracy of conclusions about the relationship between *X* and *Y* when controls are included, of course, depends on the correctness of the underlying assumptions about the role of control variables (e.g., that liking contaminates ratings of performance rather than being the effect of performance), a point made by numerous methodologists for decades (e.g., Burks, 1926; Fisher, 1925; Meehl, 1971; Pedhazur, 1997). The implicit and automatic assumption that including control variables is playing it safe ignores the possibility that such variables play an unexpected role that renders their control quite dangerous. Often controls consist of demographic characteristics and other variables considered fixed in the individual such as personality. It is assumed that such characteristics are properties that cannot have been affected by the other variables in the study. After all, one's attitudes about the job or one's job performance cannot influence one's gender. Although one's job attitude cannot (so far as we know) change one's gender, the assumption that job attitudes cannot explain gender composition in a research study does not necessarily follow. Meehl (1971) stated that making such assumptions is the single biggest vice in social science.

Meehl (1971) was quite critical of the practice of automatically including controls and discussed as an example the routine control of socioeconomic status (SES) in studies of schizophrenia. He provided evidence for six feasible mechanisms involving SES that would make its control inappropriate, leading to erroneous conclusions. Similarly, Spector et al. (2000) criticized the routine practice of controlling trait negative affectivity (NA) in studies of job stress. Using arguments similar to Meehl's (1971), they suggested six alternative mechanisms by which trait NA might relate to job stressors and strains, each of which had supporting empirical evidence for its feasibility. The fact that such alternatives have been supported makes the assumption that NA, or SES in Meehl's case, is causing distorted correlations foolish at best, leading to less rather than more accurate estimation of underlying relationships. Instead of purifying observed relationships, statistical control in such cases is muddying them, that is, leading to incorrect conclusions. The possibility and in many cases the likelihood that control variables are related to the variables of interest through a variety of mechanisms is the reason that the purification principle is a methodological urban legend.

The Vice of Concluding Demographics Cannot Be Effects

There seems to be an implicit belief that a fixed characteristic of people can be a cause but not an effect, or as Meehl (1971) noted, must be on the "input side." After all, environmental (e.g.,

workload) and relatively transient psychological (e.g., job attitudes) factors cannot affect properties like age and gender or even enduring characteristics like personality (at least in adults). Whereas this is obviously true for individual people's demographic characteristics, for example, liking one's job cannot make one older or change one's gender, liking can affect the age or gender distribution of a sample.

Take the case of gender and job satisfaction. Suppose for the sake of argument that there are gender differences in a study with women being more satisfied on average than are men on average. If we code gender as 1 = *male* and 2 = *female*, we can compute a correlation between gender and job satisfaction, the latter being assessed with a continuous scale. The difference between genders will be reflected in a positive correlation between the two variables. Although it is possible in this case for gender to have "caused" job satisfaction, for example, perhaps women are biologically prone to be happier in life and therefore their jobs, the mere existence of gender differences does not automatically obviate alternative explanations for mean differences between our two groups. One such possibility is that men and women are equally likely to be satisfied with a particular job, but women are less tolerant of dissatisfying job conditions, and when faced with a displeasurable job, they quit. Evidence suggests that on average, men may focus more time and attention on career, for example, they tend to work more hours (Hill, 2005; Spector et al., 2004) and are less prone to allow family issues to interfere with jobs (Hill, 2005). Therefore, they might be more tolerant of suboptimal job conditions and will be less likely than women to quit a job before finding a replacement that is equivalent in pay and benefits. It is not our intent to suggest that this theory is correct but merely point out that a feasible alternative is simple to generate.

If in fact dissatisfied women are more likely than men to quit their jobs in a given situation, we would expect to find the female mean to be higher than the male mean, thus to observe a relationship between gender and job satisfaction. This is illustrated in Figure 4, which shows two hypothetical distributions of job satisfaction scores, grouped by gender. Note that there are no females who are low on job satisfaction, as they have quit. The same is not true for the men. In a real sense, gender, or more correctly the gender distribution differences in these samples, is the effect rather than the cause of job satisfaction. More precisely, what is going on here is that gender moderates the relationship between job satisfaction and turnover. It is not that gender affects satisfaction (women are not inherently more satisfied than men) but rather gender influences responses to dissatisfaction, and through that mechanism, women have higher satisfaction on average than do men. Thus, one should not in this case blindly try to purify the relationship between job satisfaction and another variable by controlling gender under the implicit presumption that it is somehow inflating relationships involving job satisfaction, as to do so will lead to a biased estimate of the relations between variables and possibly to a mistaken conclusion.

Of course, one can easily derive additional feasible explanations for why men and women would differ in their job satisfaction that does not imply the impact of gender on job attitudes. Such alternatives can be generated in either direction. A mechanism favoring women would be:

Women are given better job assignments than men. It might be that the dirtiest, nastiest, and most disagreeable tasks are given to men. Thus, men on average have lower job satisfaction not because they are inherently less satisfied with work but because they are disproportionately assigned to disagreeable tasks. If one were to control gender in a study of task characteristics relating to job satisfaction, one would remove the effect of interest from the effect of interest. A mechanism favoring men would be:

Women receive more discriminatory treatment from their supervisors. This results in fewer tangible (e.g., pay and promotion) and intangible (influence and status) rewards. If one were to control gender in the study of rewards relating to job satisfaction, given this mechanism is at play, one would again be removing the relationship between the variables of interest from the relationship between the variables of interest.

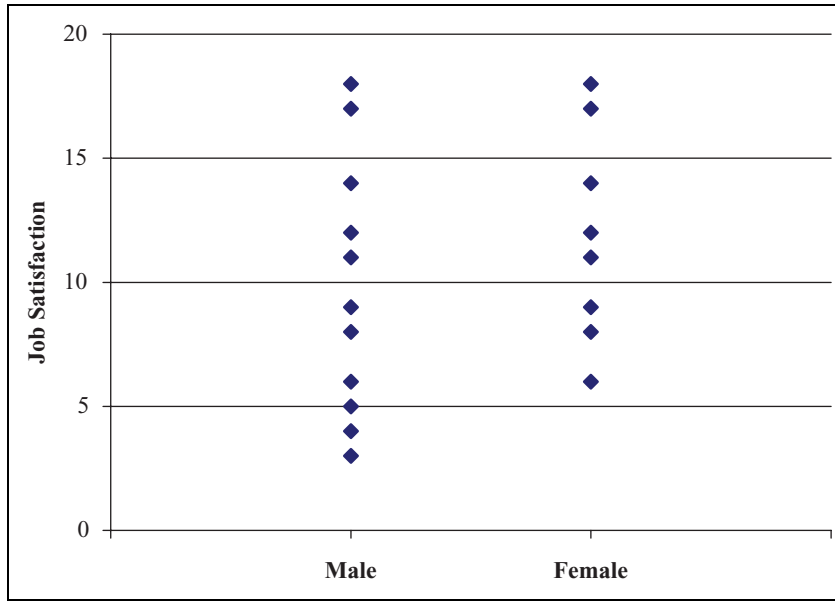

Figure 4. Hypothetical distribution of male and female job satisfaction

How One Should Use Control Variables

Throughout the literature on control variables, there seems to be more discussion about what not to do than what to do. Meehl (1971), for example, noted that he had little advice to provide to researchers, other than at least to report results with and without controls. Faced with concerns by methodologists who offer few specific solutions, many researchers adopt a "do the best you can" approach and presume that adding controls is conservative and likely to lead to a conclusion that is at least closer to the truth than omitting them. As Meehl (1971) notes, this practice is far from conservative. In fact it is in many cases quite reckless.

In the remainder of this article, we argue that the problem is not that control variables are used but rather that they are used inappropriately. The widespread assumption of the purification principle—that relationships with control variables are closer to the truth than without control variables—is unfortunate in that it has allowed for a mismatch between hypotheses and analyses and less than thorough interpretations of results. Furthermore, it has allowed researchers to avoid clearly thinking through the likely roles for all of the variables in their studies. We make two specific recommendations. First, researchers should be explicit about the hypothesized role for all variables in an analysis and have evidence upon which to base their suppositions. Controls should not be entered blindly in analyses under the belief that they will purify results. In conditions where there is evidence for contamination, it may make sense to use partialling or semipartialling to estimate the relationship between the variables of interest while controlling for the hypothesized contaminant. Other mechanisms, such as spuriousness, require additional analyses to more completely interpret results. It should be kept in mind, however, that merely stating a theory or pointing to observed relationships among control and substantive variables in the past is insufficient. A reasonable case should be made for the proposed role of control variables. If one theorizes that contamination has occurred, one should have reasonable evidence to support such a supposition.

Second, we echo Meehl's (1971) concern about the misuse of demographic variables, but go a step farther in suggesting that we rethink the use of demographics in the first place. In most cases, demographic variables serve as proxies for the real variables of interest (Breaugh, 2008). It would be more informative to focus our research attention on those variables of real theoretical interest. Attention to demographics should focus on mechanisms that explain relations with demographics rather than on the demographics themselves, which in many cases are used with little apparent concern for the reasons that demographics might relate to variables of interest. Here, we are not suggesting that demographic variables are unworthy of research. Rather, we are suggesting that they be avoided as mere control variables in theory development and testing.

Specify Roles for All Variables

As we have noted, there can be many reasons that one might find an observed pattern of results when control variables are included in an analysis. The blind inclusion of controls is unlikely to purify the results. What is needed is a more reasoned approach to the inclusion of potential control variables. If a researcher has a reasonable empirical/theoretical basis for assuming that certain "control" variables have a particular connection to other variables in the study, such reasons should be made explicit. Thus, controls should be promoted to have equal status with the other variables in the study, even though they might not be the main focus of a study. Why should personality variables, for example, be treated as second-class citizens in studies, if there is reason to believe they have important effects on the major outcomes of interest, whether the role is to affect underlying constructs of interest or contaminate their measurement? In such cases, the complete theoretical basis for the study should include the control variables. Analyses conducted in these cases should match the proposed connections among the variables and the specific hypotheses to be tested. They might involve a number of statistical techniques from simple to complex, including correlation, multiple regression, or SEM.

Testing alternative hypotheses about control variables. In most cases in which control variables are used, the researcher is trying to rule out the possibility that observed relationships of interest are due to the action of the control variables. Although not typically approached in this way, the control variables are central to implicit and unstated alternative hypotheses for results. Thus, a study's hypothesis might be that X is related to Y, based on a theory of how X and Y are connected (e.g., job satisfaction leads to turnover intentions). The unstated hypothesis is that demographic variables account for the relationship between X and Y. In many cases, the exact mechanisms by which controls might operate is unstated, but as the earlier quotes we noted suggest, researchers often have some ideas that control variables are having effects on the variables of interest. Of course, cross-sectional nonexperimental data can only provide insights into relationships among variables and not whether those relationships are causal.

The use of control variables would be far more productive, if approached as alternative hypothesis tests. Rather than just throwing control variables into analyses, typically in a first step, one should do comparative tests with and without controls to show whether their addition has an effect on observed relationships among the substantive variables of interest to the study. The ability to rule out the possibility that observed relationships were due to variables extraneous to the focal theory and hypotheses being tested is important.

As well described by Shadish, Cook, and Campbell (2002), the building of a convincing case for a causal relationship between two variables involves three steps: establishing that the two variables are related, demonstrating a temporal sequence by which the presumed cause precedes the presumed effect, and ruling out feasible alternative explanations. Cross-sectional nonexperimental designs can provide important evidence for the existence of relationships that are theoretically expected (Step 1). Furthermore, they can help rule out some feasible alternative explanations that such observed relationships are due to control variables by comparing results with and without control variables

(Step 3). Of course, this only rules out the possible effects of the control variables chosen and does not rule out other possible alternatives. Thus, the major value of the alternative hypothesis approach is to demonstrate that the effects of control variables on relationships of interest are unlikely.

To establish a convincing case that the control variable played the hypothesized role is a far more difficult undertaking that will require additional evidence and strategies (e.g., see Holland, 1988; Robins & Greenland, 1992). Making a convincing case for a causal relationship that a theory might specify requires far more than just establishing relationships and ruling out the possible effects of some control variables. However, establishing relationships and ruling out potential control variables as explanations for those relationships is a reasonable first step before other more costly and difficult studies are conducted. Tests of alternative hypotheses involving control variables can be conducted with a variety of analytic methods. A simple approach is through the use of multiple regression. The idea of using multiple regression to test alternative hypotheses is certainly not new. Hierarchical approaches are available whereby one enters a predictor variable X_1 at Step 1 and then adds an additional predictor X_2 at Step 2 to see if one can attribute the relationship of X_1 with the criterion to X_2. Of course, this logic can be expanded to many more variables. An example of this approach is White and Spector (1987), who were interested in explaining why age related to job satisfaction. They established in Step 1 that age and satisfaction were related, and then entered at Step 2 a set of variables that might theoretically account for the age effect, such as higher salaries for older workers. Their results showed that most, but not all, of the age relationship could be accounted by these other variables, leading them to conclude that they might be able to explain why older workers are more satisfied, at least in part. Note that in this case, the demographic variable was entered first, because it was the focal variable of the study and not thrown in the analysis blindly as a control. Williams and Anderson (1994) provide an SEM example of a similar approach, although their purpose was to demonstrate the potential impact of method variance (a contaminant) due to the personality trait of NA on a structural model. They compared models with and without the proposed contaminating factor and showed that it had little effect on the structural model of interest.

Researchers can approach nonexperimental studies in a hierarchical way by first generating a set of baseline hypotheses concerning variables of interest. Alternative explanations and competing hypotheses are next generated. Finally, a series of analyses would be conducted to rule in or out alternative possibilities.

Baseline hypotheses are the usual hypotheses generated in most articles today, where relationships among variables of interest are based on prior research and theory. Typically, a series of hypotheses are generated concerning relationships among different pairs of variables, with often more complex triplets involving mediation or moderation also included. Sometimes entire models are hypothesized involving a system of variables. However, rarely do researchers generate competing hypotheses, pitting one mechanism against another in an attempt to rule in or out alternative possibilities (Williams & Anderson, 1994 is a notable exception).

Alternative hypotheses would specify the role of what are generally considered control variables. These variables would be posited to "explain" the relationships among the variables of main interest. For example, it has been well established in the literature that reports of job stressors and job strains are related, and theories suggest the former leads to the latter (e.g., Spector, Dwyer, & Jex, 1988). A feasible alternative has been suggested, however, that the personality variable of NA might contaminate and inflate observed relationships (Watson, Pennebaker, & Folger, 1986). Individuals who are high in NA might be predisposed to complain about conditions in life, as well as report negative feelings and physical discomforts. Thus, this predisposition to complain would explain why individuals reporting high stressors also report high strains. A comparison of results including and excluding neuroticism from stressor–strain relationships has been used to address this issue, showing that in some cases most of the observed stressor–strain relationships could be accounted for by NA, whereas in others NA could be ruled out (Spector et al., 2000). It is important to note that in cases

where NA remained a viable alternative explanation, the mechanisms that produced the reduction in observed relationships were not clear, and in fact Spector et al. (2000) provided evidence for several mechanisms that refuted the contamination hypothesis. For example, it was shown that in some cases NA related to objective work stressors, suggesting that those high in this personality variable experienced greater levels of objective stressors on the job. They argued that controlling for NA in such studies would likely lead to erroneous inference.

Alternative hypotheses to test should be generated in the planning of a study so that appropriate measures and design features would be included. It is also possible to rule in or out variables that are not hypothesized, in an exploratory way, if the necessary variables were included in the study. In such cases, it is best to base analyses on some firm theoretical ideas and avoid shotgun approaches that capitalize on chance and inflate Type I error rates.

An important consideration in the design of studies intended to test for possible causal connections is the unmeasured variables problem (James, 1980). Analyses that omit causal variables that are related to other causal variables are likely to yield inaccurate results (however, see Mauro, 1990; Meade, Behrend, & Lance, 2009). This can occur because the estimated relationship between a predictor and criterion is really due to another predictor. As recommended by James (1980), one should do a thorough conceptual analysis of all possible related causes of a phenomenon of interest when designing studies. We recommend that such conceptual analyses include the role of demographics and other potential control variables rather than treating them as extra variables that are automatically included without firm basis.

Controlling for contamination. If one wishes to control for contamination, it is important to first be explicit about whether the contaminating variable (C) affects X, Y, or both. Statistical controls can then be used to indicate the relationship between X and Y while removing the effects or holding constant the proposed contaminating variable C. Partial correlation can be used to control for C in both X and Y. Semipartial correlation allows for the removal of C from either X or Y. It should be noted that multiple regression controls the effect of C from the X (predictor) side of the equation and does not control the potential effect of C on the Y (criterion) side. Partialling and semipartialling can be expanded to the case where there are two or more proposed contaminants. Higher order partial and semipartial correlations can be computed to show what the relationship is between variables of interest, while removing effects of multiple controls.

The possibility of contamination represents a form of alternative hypothesis that should be tested in a comparative way. Results with and without the potential contaminating variables can be contrasted in an attempt to rule in or rule out the possibility that results can be attributed to the control variables. As with any use of controls, ruling out the alternative hypothesis is far easier than ruling it in.

An example. We will illustrate our approach with data from 146 support personnel from the University of South Florida, who completed measures of workload, physical health symptoms, and state anxiety at work as a measure of mood (data are from Spector et al., 1988). Based on job stress theory, our primary Hypothesis 1 is that workload would relate to symptoms. Individuals exposed to the demands of heavy workloads would experience higher levels of symptoms, a form of physical strain. An alternative hypothesis is that the relationship between workload and symptoms is spuriously due to mood as reflected by state anxiety. Individuals who are high in anxiety might have difficulty maintaining attention, and thus they perform poorly. The inability to maintain effective effort will result in work piling up and a heightened workload. At the same time, anxiety would lead to short-term physiological responses (e.g., secretion of catecholamines and cortisol) and more long-term physical symptoms such as headache and stomach distress. Thus, alternative Hypothesis 2a is that mood is a spurious cause of the relationship between workload and symptoms. A second alternative (Hypothesis 2b) is that mood is a contaminant of the relationship between workload and symptoms. It has been suggested that NA can serve a contaminating role in the assessment of physical

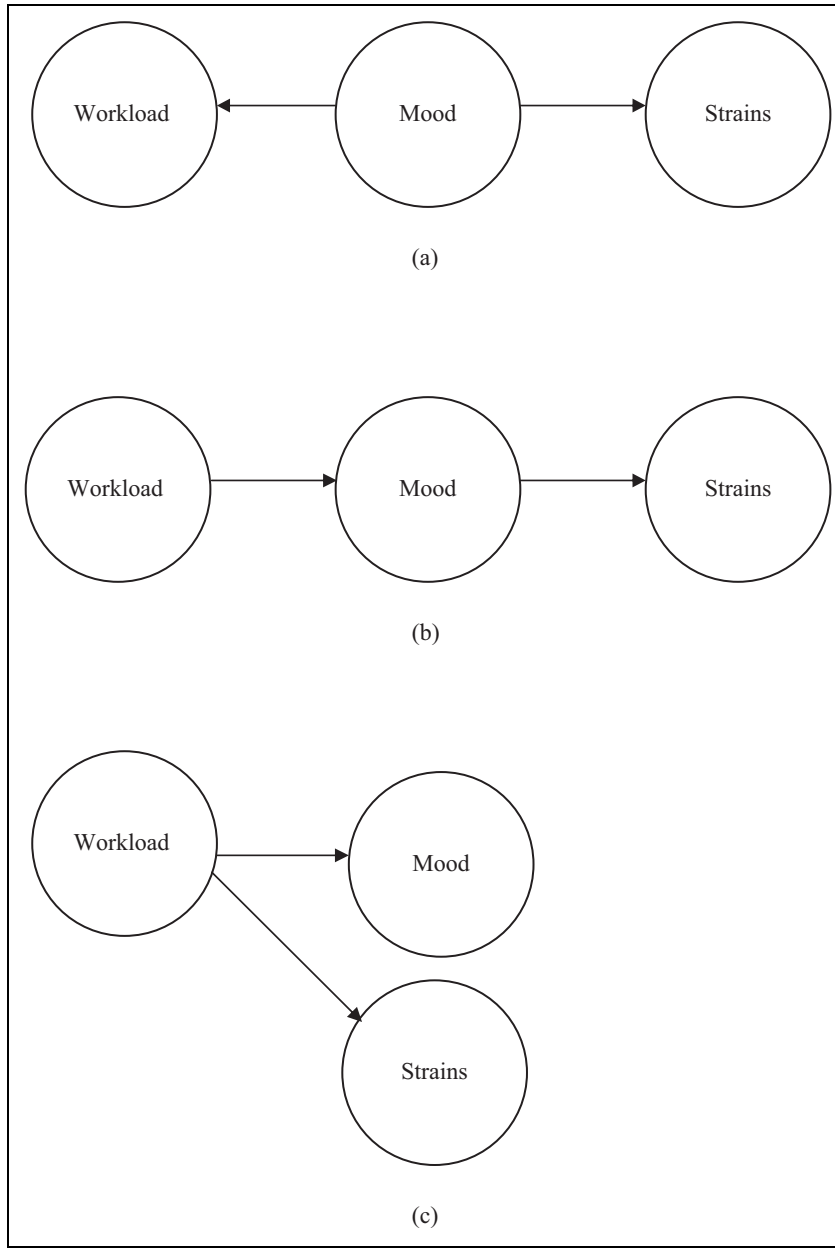

Figure 5. Alternative models relating workload, mood, and strains

symptoms (Watson & Pennebaker, 1989) as well as stressors at work (Watson et al., 1986). Individuals who are in bad moods (high-state anxiety) will tend to overreport negative experiences (heavy workloads) and internal states (physical symptoms). Thus, the assessment of workload and/or symptoms might be contaminated by mood.

A test of Hypothesis 1 is easily accomplished with the zero-order correlation between workload and symptoms. In this case, it is a significant .26, showing that employees who report higher levels of workload report more symptoms. Hypothesis 2a can be tested by comparing multiple regression

of symptoms (Y) on workload (X) with the regression of symptoms on both workload and mood (C). In the former case, the standardized coefficient for workload was a significant .24. In the second case, it reduced to a nonsignificant .09. Thus, these results support the possibility that mood was a spurious cause of both workload and symptoms. It should be kept in mind that other possible roles for mood would yield the same pattern of results, such as contamination, mediation, or that workload and symptoms caused mood rather than the reverse (see Spector et al., 2000 for several additional feasible mechanisms).

Hypothesis 2b is that mood acts as a contaminant. If this is what we believe, we would want to determine the correlation between workload and symptoms with the effects of mood removed. If we believe the contamination affects both workload and symptoms, we would compute the partial correlation, which in this case is .093. If we believe only workload is contaminated, the semipartial removing mood from workload is .087. Removing the effect of mood from only symptoms yields a correlation of .081. Our conclusion in this case is that mood could not be ruled out as a contaminant, although these results alone are inconclusive. This is still an important insight in that it tells us that conclusions concerning the impact of workload on symptoms might be premature and that we need additional work with more conclusive designs to further explore why these two variables are related. One such piece of evidence is a further finding from Spector et al. (1988) that the relationship between symptoms and supervisor-rated workload was only .07, which is significantly lower than the corresponding correlation with self-reported workload, based on a t-test for dependent correlations. Furthermore, there was reasonably high convergence between the two sources on workload ($r = .49$) suggesting they are likely assessing largely the same construct. Taken together, these findings suggest that there is something more complex going on than the simple idea that workload leads to symptoms.

How to treat demographic variables. In most investigations where they are used, demographics are proxy variables (Breaugh, 2008) and not direct measures of our variables of interest. For example, suppose we conduct a study in which age is positively related to job performance. We can propose that age is the precursor, but it is not age itself that is affecting performance, but rather factors that are associated with age. With age comes experience and job knowledge, and it may well be these variables that are the factors at play here. In other words, the demographic is rarely the object of psychological interest. It may be more productive from a theory development sense for us to focus attention on what we believe are the real mechanisms at work and include variables in our studies that address them. Assessing both the demographic and the potential mechanisms can yield valuable information about why people differ according to demographics. For example, a study in which age, experience, and knowledge are all included will allow for tests of hypotheses about why age and performance are related. This is the approach taken by White and Spector (1987) who included several factors they thought would explain why older workers were more satisfied than younger workers.

A thorough conceptual/theoretical analysis of what the role of demographics might be will likely inform the design of studies, which will allow for more complete investigations of the phenomena of interest. If such variables are thought to be important, they are deserving of just as much attention as the main variables of focus in our studies. If it is believed that such variables are unimportant and have little relationship with variables in a particular domain, then they do not need to be controlled or investigated. It is time to get beyond blindly controlling demographics in a vain attempt to purify our results from them.

New Thinking About Control Variables

Much of the organizational research literature, as well as the literatures of some other disciplines that rely on field research methods, can be quite disjointed and nonsystematic. Researchers tend to focus

on establishing relationships among large numbers of variables rather than attempting to understand the underlying processes involved. Much of this tendency is undoubtedly driven by the relative ease in establishing relationships and difficulties in establishing processes, but much of it is also due to the norms that have arisen in how research is done. Thus, an individual might devise a new construct and scale to assess it in one study and then conducts a series of subsequent studies designed to determine the nomological network of relationships with a large variety of variables that should be related in theory. The norms strongly encourage including multiple control variables in analyses designed to explore relationships among variables of interest.

An alternative approach would be instead to establish a single relationship of the new construct with another variable (or small number of variables) and then systematically attempt to explain the reasons for that relationship through a series of alternative hypothesis tests. Returning to our gender and job satisfaction example, one might first ascertain if there are gender differences, and then conduct a series of studies and analyses to test alternative explanations. Initial tests might be done with simple cross-sectional survey methods, with promising leads followed up on with other more definitive, although more difficult, methods. For example, suppose we find that women report better interpersonal treatment at work than do men, and this explains the greater satisfaction of women. We could do additional research to independently confirm that this is or is not the case. One approach would be to have observers rate the treatment men and women receive, either through live observation or by viewing video recordings taken in a workplace. If subjects of the study are customer service employees, and the poor treatment comes from customers, one could make video recordings of customer interactions that are coded by judges for nastiness. To control for knowledge of whether the target employee is male or female, the recordings might show only the customer. Audio recordings might include only the customer's voice.

Although our advice might seem to be directed mainly to authors of papers, it is relevant to editors and reviewers of papers as well for two reasons. First, editors and reviewers should be wary of the inappropriate use of control variables in papers they handle. At the very least, authors should be asked to thoroughly explain and justify what they have done. They should not get away with merely saying they included a control variable just because it might affect the variables of interest. Furthermore, there should be full disclosure about the effects of adding the control variables, with complete discussion of any discrepancies. Editors and reviewers should be the first line of defense against the purification principle. Second, in our experience, it is the review process that is often responsible for the blind use of control variables, as authors are asked to include controls in their hypothesis tests, even though the control is not relevant to the underlying theoretical framework that drove the study. Editors and reviewers should not insist that authors add controls to their analyses unless there is a solid justification for doing so, and if they are included, they should be approached as alternative hypothesis tests. It should be incumbent on editors and reviewers to explain why their requested control variable is important to include. In other words, editors and reviewers should be just as judicious in their requests for controls as are authors in their initial inclusion of controls.

Other recent authors have offered advice about how control variables should be handled in organizational research. Becker (2005) listed a dozen recommendations, including how control variables are selected, measured, reported, and interpreted, which are quite compatible with our advice. We go beyond Becker by focusing attention on the distinction between contamination, spuriousness, and other roles, by providing specific strategies for comparative model tests, and by discussing better uses of demographic variables.

Our examples have been from the micro-organizational realm, but the issues we discuss hold equally for macro-organizational research where even more control variables are used routinely (Atinc & Simmering, 2008). Many of the control variables used in such studies are relatively objective characteristics of firms, such as financial indices or size, which might be considered the organization-level version of person-level demographics. The same logic applies, however, in that

there can be many reasons that control variables might relate to the variables of interest in a study, making it possible that statistical control would distort observed relationships and lead to erroneous inference. We might assume, for example, that firm size might distort relationships between HR practice and financial performance, but it is possible that size is the result of practices and performance, such that controlling for size is statistically removing the very effects one is wishing to study.

As organizational scientists, we have at our disposal a wide variety of statistical tools that can help us draw inferences about the variables in our studies. Unfortunately, all too often we fail to take full advantage of those tools, resulting in missed opportunities at best, and erroneous inference at worst. We call for a change in how we think about statistical control variables. It is time to give our "control" variables the attention they deserve. Rather than being throw away, variables that are automatically and blindly added to our analyses in the hope that they will purify our results, we should give them their due and make them first-class citizens in our investigations. A systematic investigation of variables routinely used as controls will go a long way toward advancing our understanding of why the variables we study might be related.

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

Atinc, G., & Simmering, M. J. (2008, October 29–November 1). *Use of control variables in management research*. Paper presented at the Southern Management Association, St Petersburg Beach, FL.

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*, 274-289.

Bidwell, M. J., & Briscoe, F. (2009). Who contracts? Determinants of the decision to work as an independent contractor among information technology workers. *Academy of Management Journal*, *52*, 1148-1168.

Breaugh, J. A. (2008). Important considerations in using statistical procedures to control for nuisance variables in non-experimental studies. *Human Resource Management Review*, *18*, 282-293.

Burks, B. (1926). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology*, *17*, 625-630.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cuervo-Cazurra, A., & Dau, L. A. (2009). Promarket reforms and firm profitability in developing countries. *Academy of Management Journal*, *52*, 1348-1368.

Fisher, R. A. (1925). *Statistical methods for research workers.* Edinburgh, England: Oliver and Boyd.

He, J., & Wang, H. C. (2009). Innovative knowledge assets and economic performance: The asymmetric roles of incentives and monitoring. *Academy of Management Journal*, *52*, 919-938.

Hill, E. (2005). Work-family facilitation and conflict, working fathers and mothers, work-family stressors and support. *Journal of Family Issues*, *26*, 793-819.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, *18*, 449-484.

Hoobler, J. M., Wayne, S. J., & Lemmon, G. (2009). Bosses' perceptions of family-work conflict and women's promotability: Glass ceiling effects. *Academy of Management Journal*, *52*, 939-957.

James, L. R. (1980). The unmeasured variables problem in path analysis. *Journal of Applied Psychology*, *65*, 415-421.

Kennedy, M. T., & Fiss, P. C. (2009). Institutionalization, framing, and diffusion: The logic of TQM adoption and implementation decisions among U.S. hospitals. *Academy of Management Journal*, *52*, 897-918.

Leana, C., Appelbaum, E., & Shevchuk, I. (2009). Work process and quality of care in early childhood education: The role of job crafting. *Academy of Management Journal*, *52*, 1169-1192.

Lee, B. H. (2009). The infrastructure of collective action and policy content diffusion in the organic food industry. *Academy of Management Journal*, *52*, 1247-1269.

Lowin, A., & Craig, J. R. (1968). The influence of level of performance on managerial style: An experimental object-lesson in the ambiguity of correlational data. *Organizational Behavior & Human Performance*, *3*, 440-458.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*, 173-181.

Madsen, P. M. (2009). Does corporate investment drive a "race to the bottom" in environmental protection? A reexamination of the effect of environmental regulation on investment. *Academy of Management Journal*, *52*, 1297-1318.

Marquis, C., & Huang, Z. H. I. (2009). The contingent nature of public policy and the growth of U.S. commercial banking. *Academy of Management Journal*, *52*, 1222-1246.

Mauro, R. (1990). Understanding L.O.V.E. (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin*, *108*, 314-329.

McDermott, G. A., Corredoira, R. A., & Kruse, G. (2009). Public-private institutions as catalysts of upgrading in emerging market societies. *Academy of Management Journal*, *52*, 1270-1296.

Meade, A. W., Behrend, T. S., & Lance, C. E. (2009). Dr. StrangeLOVE, or: How I learned to stop worrying and love omitted variables. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 89-106). New York, NY: Routledge/Taylor & Francis Group.

Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science. Vol. IV. Analyses of theories and methods of physics and psychology* (pp. 372-402). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*, *77*, 143-148.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York, NY: McGraw Hill.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Pil, F. K., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *Academy of Management Journal*, *52*, 1101-1124.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*, 143-155.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shaw, J. D., Dineen, B. R., Fang, R., & Vellella, R. F. (2009). Employee-organization exchange relationships, HRM practices, and quit rates of good and poor performers. *Academy of Management Journal*, *52*, 1016-1033.

Spector, P. E., Cooper, C. L., Poelmans, S., Allen, T. D., O'Driscoll, M., Sanchez, J. I., ... Yu, S. (2004). A cross-national comparative study of work-family stressors, working hours, and well-being: China and Latin America versus the Anglo world. *Personnel Psychology*, *57*, 119-142.

Spector, P. E., Dwyer, D. J., & Jex, S. M. (1988). Relation of job stressors to affective, health, and performance outcomes: A comparison of multiple data sources. *Journal of Applied Psychology*, *73*, 11-19.

Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, *21*, 79-95.

Tzabbar, D. (2009). When does scientist recruitment affect technological repositioning? *Academy of Management Journal*, *52*, 873-896.

Varma, A., Denisi, A. S., & Peters, L. H. (1996). Interpersonal affect and performance appraisal: A field study. *Personnel Psychology*, *49*, 341-360.

Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review*, *96*, 234-254.

Watson, D., Pennebaker, J. W., & Folger, R. (1986). Beyond negative affectivity: Measuring stress and satisfaction in the workplace. *Journal of Organizational Behavior Management*, *8*, 141-157.

White, A. T., & Spector, P. E. (1987). An investigation of age-related factors in the age-job-satisfaction relationship. *Psychology and Aging*, *2*, 261-265.

Williams, L. J., & Anderson, S. E. (1994). An alternative approach to method effects by using latent-variable models: Applications in organizational behavior research. *Journal of Applied Psychology*, *79*, 323-331.

Bios

Paul E. Spector is a distinguished university professor of industrial/organizational (I/O) psychology and I/O doctoral program director at the University of South Florida. He is also director of the NIOSH funded Sunshine Education and Research Center's Occupational Health Psychology program. He is the Associate Editor for Point/Counterpoint for *Journal of Organizational Behavior*, and Associate Editor for *Work & Stress*, is on the editorial boards of *Journal of Applied Psychology, Organizational Research Methods*, and *Personnel Psychology*.

Michael T. Brannick is a professor and Interim Chair in the Psychology Department at the University of South Florida. He received his PhD in industrial and organizational psychology from Bowling Green State University in 1986. He is the author of numerous articles on research methods and currently serves on the editorial boards of Journal of Applied Psychology and Human Performance.