A Typology of Outcome Patterns in Three-Variable Models: The Pervasive Role of Mediation in Causal Systems

Gary F. Koeske
University of Pittsburgh

Randi Koeske
University of Pittsburgh at Greensburg

Details of published version, with errors:


Correspondence concerning this article should be addressed to Gary F. Koeske, Ph.D., School of Social Work, University of Pittsburgh, Pittsburgh, PA 15260.

ANNOTATIONS: Paul van Schaik, PhD

Word count: 7,342 (Abstract through References)
Abstract

Developing a lexicon of variable types is essential for effective research, but existing variable typologies use different terms for the same process, and remain incomplete. We offer a typology in which extraneousness is distinguished from mediation and moderation; suppression is subsumed under mediation; and the hypothesized role of X or Z in a three-variable system can be recast. Simple multiple regression and path analytic examples are used to clarify the difference between extraneousness, moderation, and mediation, and suppressive mediation. The practical and theoretical importance of mediation, including a rule-of-thumb for identifying its relevance, is addressed.
A Typology of Outcome Patterns in Three-Variable Models: The Pervasive Role of Mediation in Causal Systems

1 Introduction

One of the earliest concepts we teach to fledgling researchers is extraneousness or spuriousness. This is usually done in the context of causal analysis and introduces the learner to the world of three-variable models, in which third variables (Z) can participate in the relationship between X, a presumed independent (and potentially causal) variable, and Y, a presumed dependent (outcome, criterion) variable.

An understanding of three-variable relationships is foundational in mainstream research on variable interrelationships, and has a long history in psychology and sociology – the two disciplines on which most social work researchers have based their own research training. These two disciplines have evolved somewhat different variable lexicons and broad methodological languages, however. In the half century during which three-variable language systems have been discussed, relative uniformity in the use of variable terms has been achieved in the technical journal literature, but inconsistent usage and occasional confusion remain. There is also no contemporary synthesis of similar or overlapping terms.

We believe the absence of a uniform variable lexicon for use in social, behavioral, and social work research constitutes more than an annoyance, because inconsistent and confused terminology sometimes reflects or
contributes to confused thinking, which has the potential to generate problematic analysis or interpretive conclusions. The purpose of this paper is to clarify the key terminology and concepts needed to describe the operation of multivariate systems, particularly those made up of three variables. Our focus will be on extraneousness, mediation, and suppression in causal analysis, with secondary reference to moderation. We propose the term recast mediation for situations in which the spuriousness of an X—Y relationship is disconfirmed by controlling for Z, and also differentiate three types of suppression that may characterize three-variable systems. Although not part of many classical presentations on causal inference, recast mediation and suppression occur often enough in actual research to warrant their inclusion in the present discussion.

Variable Terminology in Sociology, Psychology, and Social Work

Before defining and differentiating several new variable terms and proposing a terminology synthesis, we offer a brief overview of the current state of variable terminology. Within sociology, the classical methodological work on the “elaboration model” (Lazarsfeld & Rosenberg, 1955; Rosenberg, 1968) has provided a variable language that is highly workable for researchers who study small causal systems and rely on correlational designs and category measurement. Data analysis within this tradition typically involves tabular analysis, in which bivariate (X—Y) relationships are analyzed first for all cases and then broken down into “partial tables” (in which a Z is controlled). The partial tables are developed by disaggregating the data to examine the effect of one “third variable” at a time.
The most comprehensive examples of tabular analysis were presented in methodological classics by Hirschi & Selvin (1973) and Rosenberg (1968). In recent years, simple bivariate crosstabulations have given way to multi-way contingency tables, and log linear analysis (for variables measured at the category level). The work of Davis (1971; 1985) provides an important example of related work intermediate between the earliest tabular approaches and current multivariate strategies.

*Extraneous or confounded variables/threats to internal validity.* Within this framework, one role a third variable can play is that of *extraneous variable*, whose uncontrolled influence produces an apparent X—Y relationship that is actually *spurious.* An extraneous variable is one that is (typically) temporally or logically prior to both X and Y, and substantially correlated (“confounded”) with both variables. When it is statistically controlled (by examining the X—Y relationship separately for discrete levels of Z in a partial table), the X—Y relationship “disappears” or approaches zero. Extraneousness is thus the antithesis of causality, in that Y is not caused by X, the hypothesized independent variable, but by Z, the extraneous third variable. The apparent relationship between X and Y is also “due to Z”, in that Z influences both X and Y. This definition and these terms have persisted to the present day, and most introductory research methods textbooks accurately present as foundational the centrality of correctly interpreting a proposed X—Y relationship.

Within the discipline of psychology, the term *extraneous* is not often mentioned, due to the frequent use of experimental methodology and random
Mediation in Causal Systems

assignment to conditions (i.e., random exposure to X). Extraneous variables exert a systematic effect on X and Y, but randomization is assumed to equalize the impact of all potentially extraneous third variables by allowing the researcher to hold other influences constant while controlling exposure to the independent variable, X (cf. Campbell & Stanley, 1967). In psychology, too, potentially extraneous variables are often controlled through selection criteria or matching, in which case they are usually known as control variables, and cannot be a source of plausible rival hypotheses (alternated explanations) concerning the effect of X on Y.

Of course, randomization can break down (as in treatment studies, where differential dropout from the experimental and control conditions may occur over time); variable manipulations can fail or be judged methodologically suspect; uncontrolled (subject, status, attribute) variables may be measured alongside experimentally manipulated ones; and many psychologists continue to use quasi-experimental and correlational research designs. So the concept of spuriousness is as relevant for psychologists as it is for other behaviorally and socially oriented researchers. In psychology, however, Campbell & Stanley’s analysis of quasi- and non-experimental designs has provided the primary language system and conceptual scheme, with the result that psychologists widely refer to potentially extraneous variables as threats to internal validity, and understand them as a key source of plausible rival hypotheses.

Intervening variables/mediation. In early sociological discussions (e.g., Rosenberg, 1968), several patterns of interrelationship involving third variables
could result. If a third variable followed an independent variable in time, but preceded and produced effects on a dependent variable, the process was labeled *interpretation*, and the Z variable was referred to as an *intervening* variable. This labeling was identical to that used in much of the psychological literature in which an intervening variable was a measured or measurable third variable operating between X and Y and supplying an explanation for their relationship.

The early sociological term *interpretation* (of the X—Y relationship) meant the same thing as the psychological term *explanation*, but was distinguished from *explaining away*, which was used in the elaboration lexicon to describe what happens to the X—Y relationship when Z functions as an extraneous variable. The term *interpretation* has lost vitality in current research, so that *explain* and *explanation* have become the most common ways of describing the operation of an intervening variable. Other terms currently used synonymously with intervening variable are *process variable*, *mechanism*, and *mediator*. *Mediator* seems to be the most popular term in use.

In the elaboration model, interpretation, or an intervening variable, was inferred when the X—Y relationship disappeared in the partials, i.e., at each level of Z, the intervening variable. Thus, the same tabular analytic outcome was expected for an intervening variable as for an extraneous variable. The essential difference was that Z’s influence occurred prior to (or coterminous with) X in the case of extraneousness, but subsequent to X in the case of interpretation/mediation (cf. Rosenberg, 1968). Thus conceptualization and
logical inference, not a unique pattern of statistical results, were thus needed to distinguish extraneous variables from mediating/intervening ones.

**Moderation/interaction/threats to external validity.** Some references in which mediation is differentiated from *moderation* may be found in methodological articles (e.g., Baron & Kenny, 1986). Current teachers of research methods work hard to pass on the distinction to their students. The classical sociological presentations of moderation (e.g., Hirshi & Selvin, 1973; Rosenberg, 1968) used the term *specification* to denote the operation of a third variable that acted as a moderator variable by setting conditions on the generality, strength or direction of the X—Y relationship. A Z variable acting as a moderator has also been called a *conditional variable* or a *qualifier* (Rosenberg, 1968).

Regardless of terminology, the pattern of partial table results required to infer specification is the presence of *different* relationships between X and Y across the separate levels of Z. In other words, what is true for the X—Y relationship at one level of Z is not true for another, so the X—Y relationship is not general, but depends on the value of a third variable. Unlike cases of extraneousness and mediation, the temporal position of a moderator relative to X is irrelevant to its definition. A moderator is simply another predictor that interacts with X to jointly explain Y. Its influence does not negate (as in extraneousness) or account for (as in mediation) the impact of X on Y.

Psychologists had already established a long tradition of examining statistical interaction effects in factorial experiments involving two or more
independent variables. The joint effect or *interaction* of any two (or more) variables is the same process referred to as specification in the classical sociological literature. Today, both of these disciplines and social work seem to have settled on the term *moderation* to describe this phenomenon, following Saunders' (1956) early description of moderator variables in a regression context. Campbell & Stanley's (1967) concept of *external validity*, and the well-established concept of *generalizability* refer to the same process: the relationship of X with Y is not universal when Z acts as a moderator, i.e., the X—Y relationship does not generalize across all levels of the designated third variable, Z. Another related expression is *invariant*, as in “the X—Y relationship is not invariant across Z”, used to describe moderation.

Thus, the terms *mediation* and *moderation* have come to describe two of the important causal roles a third variable may assume in connection with X, while *extraneousness* or spuriousness has come to mean that X is “not causal” in impacting Y because Z, with which it is correlated, “really” accounts for the apparent X—Y relationship.

*Suppression*. A fourth variable role that has received less attention is *suppression*. It can often be regarded as a type of mediation, but was not included in most classical discussions of mediation penned by either sociologists or psychologists. In a three-variable model, suppression occurs when no overall association of X and Y is observed, but controlling for Z (i.e., examining the X—Y relationship at different levels of Z) reveals evidence of an X—Y relationship that is relatively uniform (invariant) at each level of Z.
Rosenberg (1968) referred to suppression as that case in which an overall bivariate table shows no association between X and Y, but a uniform X—Y relationship becomes apparent in each partial table. He also referred to *distorter* variables – third variables that, when controlled, yield an X—Y relationship for each partial table that is opposite in sign to their overall bivariate relationship. We will refer to distorter variables as one form of suppression, following Tzelgov’s comprehensive technical work on suppressor variables in education and psychology (Tzelgov & Henik, 1991; Tzelgov & Stern, 1978). Suppressive phenomena may be less frequent in research practice than other three-variable systems, but they are far from unimportant and deserve special attention because their relevance to causal inference remains less widely understood.

*The need for clarity and consistency in variable terminology.* Although the professional journal literature has adopted a fairly uniform usage for the four primary variable terms (extraneous variable, mediator, moderator, and suppressor), and has characterized each by means of accurate definitions and distinctions, many introductory and intermediate textbook presentations on research and statistics have not followed suit. This is unfortunate, because new students may receive inconsistent, incomplete, or inaccurate instruction and this may impede their ability to write and think coherently as researchers or consumers of research.

A thorough knowledge of third variables amounts to an understanding of the very structure and process of mainstream relationship-testing research. Even a cursory review of this foundational content in introductory texts reveals
instances in which (a) the concepts receive minimal coverage and “intervening variable” is seemingly used to represent any third variable; (b) confused and inaccurate use is made of terms like intervening variable, mediator, or spuriousness; and (c) moderation and suppression are confused. Even a popular introductory social work research methods textbook with the most extensive explanation of variable types provides a tabular example of interpretation/ intervening variable for which the relationship in the partial tables is not close to zero. Suppression, the variable phenomenon that is least often presented and most often misunderstood, is presented accurately in the same text, however.

Brief and inaccurate presentations of variable terminology in some introductory texts may reflect a judgment that such material is of secondary importance. Our belief is that such material is actually essential to the conceptualization and conduct of quality research. Thorough coverage of variables is not too advanced or peripheral a topic for introductory texts on research methods. Both the intelligent consumption of research findings and the production of high quality research depend on a complete understanding of the few pivotal concepts covered in this paper.

2 Toward A Typology of Variable Effects in Small Variable Systems

The concept of extraneousness offers an excellent starting point for any discussion of the roles of variables. Mediators address the issue of *why* a relationship occurs, moderators tell us *when* or *where* they occur, but potentially
extraneous variables tell us if a genuine causal relationship occurs at all. Why and when are generally irrelevant until if has been established.

While it is essential for researchers interested in inferring causation to rule out plausible rival hypotheses that result from the uncontrolled influence of unrecognized or poorly conceived antecedent variables, too narrow a focus on accepting or rejecting a proposed X—Y relationship may impede understanding of larger three-variable and multivariate systems. Typically, an antecedent third variable, Z, is a second predictor of Y. We may conceptualize such a Z not as a confound, that completely negates the effect of X on Y, but as part of a three-variable system in which Y has two predictors, X and Z.

This reconceptualization makes it easier to recognize that the estimate of X’s influence may change because of the presence of Z, and that the estimate of Z’s influence may change, due to the presence of X. Such a reframing allows us to consider variable relationships in multiple regression terms, instead of in tabular analytic terms alone. It may help guard against Type II as well as Type I errors.

*Multiple regression for three- (or more) variable systems.* In most correlational research, multiple predictors, which may be correlated with one another, are used to explain the variability in the criterion, Y. The unique influence of each predictor is usually estimated by partialling out its statistical contribution to Y in a multiple regression analysis.

In the remainder of this paper, we adopt the language of regression analysis, making the simplifying assumption that predictor variables (X, Z) are
either dichotomous or represent scores (interval-level measures) on some dimension, while the criterion, Y, is a score (i.e., interval or ratio measured) variable. Using the terminology of multiple regression helps to clarify the causal importance of suppression, and is more efficient and far easier to understand than tabular analysis for variable systems of more than three variables. Parallel tabular examples could be provided in every case discussed below.

Several three-variable models are presented in Figures 1 to 2, using correlation/regression terms. To make sense of the figures, the reader needs to conceptually understand the statistical concept of partialled estimates or coefficients, such as semi-partial and partial correlation, or standardized partial regression coefficient (beta), as these terms are used in multiple regression. No equations are developed in the discussion that follows. Instead, the reader is asked only to understand that, when there is more than one predictor of Y (say, X and Z), the explanation of Y that they share is removed from (partialled out of) their respective estimates. The result is an estimate of the effect of X controlling for Z, and of Z controlling for X, in any three variable (X, Z → Y) system. Including additional predictors (Z₁, Z₂, Z₃, etc.) complicates the picture, of course, but the essential language of variable types remains.

The causal interpretative issues that correlational researchers always confront become germane for experimenters whenever randomization is lost or when non-manipulated variables are of causal interest. Researchers know that, technically, all plausible third variables must be controlled before a causal statement can be considered warranted. They also know that it is practically
impossible to discount all third variables in a correlational design. They do their best by anticipating and measuring the most compelling third-variable explanations.

The researcher may introduce a potentially extraneous third variable and find that the X—Y relationship statistically dissolves, i.e., it becomes zero or nearly zero, as in textbook examples of extraneousness. In such cases, the Z variable represents an extraneous variable and the relationship between X and Y must be viewed as spurious and non-causal.

But often, rather than dissolving altogether, the X—Y relationship is diminished by controlling for Z, yet remains statistically significant. This is the situation we will discuss most completely, relying on diagrams to convey important examples.

**Extraneousness.** Extraneousness is one of four general results that may occur when a third variable, Z, operates as an antecedent (Za) in the prediction of Y from X. Extraneousness represents only one of the four Za possibilities – the one in which X does not survive as a potential cause of Y once Z is controlled.

Figure 1 shows a series of simple 3-variable path diagrams representing the four possible outcomes for Za as a second predictor of Y. In each diagram, the number on the arrowed line prior to the slash (/) is the simple (or zero-order) correlation; the number after the slash is the standardized regression coefficient (beta or β). The number on the Za—X line is the intercorrelation or collinearity of the predictors. The patterns depicted are restricted to situations in which the Z
variable has a non-zero correlation with X, the X—Y and Z—Y correlations are not both zero, and Z is either an antecedent or a consequent of X.\(^1\). The diagram of Case A in the figure represents extraneousness, while the other three sets of diagrams (Cases B, C and D) represent various forms of mediation by X that were unanticipated in the original research hypothesis, which concerned X as the independent variable.

The coefficients shown in each set of diagrams represent hypothetical but concrete results that distinguish the four possible alternatives. The path coefficients (betas) were obtained by formula from the hypothetical simple correlations, partial correlations, semi-partial correlations, and the predictor tolerances. If we imagine that about 100 cases are involved, a correlation of about .20 would be significant at the .05 level. A relationship of the magnitude depicted would probably also represent a meaningful effect size (i.e., be of substantive significance). In each example, Z—X and X—Y relationships of a particular sign are shown. Examples of opposite sign could be added, but they would not represent different processes, and would simply be mirror images of the examples depicted.

In the diagram for Case A, *Extraneousness*, the initially significant X—Y relationship approaches zero and is not significant in the three-variable model (i.e., as a beta). The initial X—Y relationship must now be regarded as spurious, i.e., X cannot be a viable cause of Y. We see that such a result requires substantial relationships between Z and both X and Y.
3 Three Patterns of Unsuspected Mediation Involving an Antecedent Third Variable, Za

Cases B through D in Figure 1 represent a phenomenon we call *recast mediation*. In each example, the original independent variable, X, has been recast as a mediator between Za and Y. This pattern of interrelationship may be more likely in actual research situations than classical extraneousness. A researcher who is focused exclusively on the predicted X—Y relationship, however, may conclude only that the X variable has survived the test of causality, since spuriousness cannot be claimed in relation to Za. But realizing that X should be tentatively considered a cause of Y is an incomplete, though accurate conclusion. It is important to attend to the entire multivariate system and realize that X has been recast into the mediator role, with Z as an antecedent independent variable.

*Redundant recast mediation.* Case B in Figure 1 shows a result of Redundant Recast Mediation, which means that (a) X has been recast as a mediator and (b) the relations between X and Z are redundant, i.e., the predictors share in their explanation of Y. In this pattern, the effect size estimates for X and Z, i.e., their betas, are smaller than their corresponding simple (bivariate) correlations. What should be apparent in this model is that the antecedent variable, Za, has both a direct effect on Y (beta = .26) and an indirect effect on Y through X (Z → X → Y). The size of this path is obtained by multiplying the path coefficients along the path [(.40 × .35 = .14)].
The redundant nature of the explanatory system is apparent for two defining reasons: (a) both betas are less than their corresponding correlations, and (b) both the direct effect of Za on Y (+ .26) and the indirect effect of Za through X (+ .14) have the same algebraic sign (in this case, positive).

**Suppressive recast mediation: three types.** A less frequent occurrence, but probably not rare, is shown in Case C, called *Suppressive Recast Mediation*. This pattern represents a result that may be thought of as the opposite of redundant explanation. In suppressive recast mediation, the X—Y relationship takes on one of three variants (for the three-variable model), rather than yielding a beta that is less than its corresponding zero-order correlation. We call the three variants *Effect Unmasking*, *Effect Intensification*, and *Effect Reversal* (cf. Koeske & Koeske, 2002)

In effect unmasking, a zero correlation between X and Y becomes significant when Z is added as a predictor to the model. Two examples of this process are shown in the diagram – one in which the effect of X is unmasked (labeled i, in which the X—Y relationship is initially zero), and one in which the effect of Z is unmasked (labeled ii, in which the Z—Y relationship is initially zero). In effect intensification, a relatively small relationship becomes noticeably larger and significant. And in effect reversal, the beta estimate in the three-variable model is opposite in sign to the simple X—Y correlation. In both effect unmasking and effect reversal, the Z—Y path coefficient (beta) is *larger* than the simple correlation, and in effect intensification, it may either become larger or change sign.
Unmasking the effect of X is of obvious importance, since a potential cause of Y could be missed if X were not recognized as part of a suppressive system. Researchers often rapidly lose interest in near-zero and non-significant correlations, but as the example labeled (i) shows, the potentially causal effect of X is “unmasked” when Za enters the model. Note should be made of the occurrence of opposite signs for the direct effect of Z (−.51) and its indirect effect through X [(.50 × .26 = +.13)].

An example often used in research books to exemplify extraneousness, the fire truck example (see Kendall & Lazarsfeld, 1950), should actually be regarded as an example of effect reversal suppression, the third type of suppressive recast mediation represented in Figure 1, Case C. In this example, the reader is shocked to find that the more fire trucks (X) sent to a fire, the larger is the ensuing damage estimate (Y). When the size of the fire (Za) is controlled, however (i.e., when Z enters the model as a second predictor), X no longer relates positively to Y, as it did in the bivariate correlation.

This result is not best understood as an example of extraneousness. Rather, it more reasonably is an example of suppressive mediation, in that we would surely hope and expect that the initially positive X—Y relationship would become negative in the three-variable model test, with coefficients much like those in the effect reversal example provided. Similarly, in the example of suppressive recast mediation shown in Figure 4 and discussed below, additional counseling sessions might predict higher subsequent symptom occurrence, but controlling for initial severity of the presenting problem (Z) should reveal the
expected benefit of additional counseling. In other words, a negative partial coefficient would be found for X in the three-variable model to which Z had been added.

Both the fire truck and clinical symptoms examples illustrate suppressive systems in operation. In effect reversal suppression, the X—Y relationship changes sign, the Z—Y beta is larger than the Z—Y correlation, and the indirect effect of Za through X is opposite in sign to Z's direct effect on Y. Opposite signs represent a defining feature of all types of suppressive mediation.

*Effect intensification suppression*, the second type of suppressive recast mediation shown in Figure 1, Case C, is also of practical and theoretical importance. It has been suggested that some of the original interpretations of the efficacy of Head Start's compensatory education programs were flawed, because the effect of Head Start, X, was underestimated (Langbein, 1980). If prior achievement level, which was lower for recipients of Head Start, had been controlled (i.e., entered as Za in the model), then the initially small estimates of the impact of Head Start might have increased. Such a relationship pattern represents suppressive mediation of the effect intensification type.

*Simple recast mediation*. The fourth possible outcome in Figure 1 is summarized under Case D. This pattern, labeled *Simple Recast Mediation*, is one for which the Z—Y relationship has diminished to near-zero in the three-variable model, resulting in only one causal path, Z→X→Y. X is recast as a mediator, as in Cases B and C, but there is no direct path from Z to Y. This pattern represents neither redundancy nor suppression.
For completeness, two examples are given for simple recast mediation, one for which an initially substantial relationship has a zero partial estimate, and a second for which an initially zero correlation of Z—Y remains non-significant (near zero) when estimated in the three-variable model. What is apparent in Cases B, C, and D (i.e., all outcomes involving an antecedent third variable, Za, except that of extraneousness, Case A) is that the model involves mediation, and the X—Y relationship is always causal within the usual limits of inferring causation in correlational research.

4 Four Patterns of Mediation Involving a Proposed Intervening Variable, Zi

The variable patterns shown in Figure 1 could be slightly modified to produce four related general patterns. These are situations in which the third variable, Zi, is proposed as an intervening (i) variable or mediator of the X—Y relationship. For these cases, the Za becomes a Zi, and the causal arrow points from X to Z rather than from Z to X. X is not recast as a mediator, as was true in Figure 1, Cases B, C, and D; and Z is instead proposed to mediate the X—Y relationship.

Classical mediation, redundant mediation, suppressive mediation, and disconfirmed mediation. The situation parallel to Case A in Figure 1 would occur when Classical Mediation is confirmed (see Figure 2, Case E): a simple correlation between X and Y “disappears” in the three-variable model, because its causal effect is exerted exclusively through Zi. Figure 2, Case F shows the case of Redundant Mediation (which is parallel to Case B in Figure 1). Case G in Figure 2 has been labeled Suppressive Mediation, and is parallel to Case C in
Figure 1. Finally, Figure 2 shows Case H, in which the proposed or expected mediation was disconfirmed, because the necessary Z—Y link was not significant. In this case of disconfirmed mediation, the X remains a potential cause of Y, but the explanation for this relationship is not Z. Thus, we have four distinguishable patterns based on the assumption that Z is an antecedent variable, discussed above and summarized in Figure 1, and four additional patterns arising when Z is proposed as an intervening or mediating variable between X and Y (summarized in Figure 2).

Summary. Of the eight possible patterns depicted in Figures 1 and 2, all but one, extraneousness, involve a potentially causal effect of X on Y; and all but two, extraneousness and disconfirmed mediation, involve a mediated effect.

5 Distinguishing Extraneousness, Moderation and Mediation

The primary purpose of this paper is to present a typology of three variable outcomes that differentiate extraneousness from multiple patterns of mediation. One type of suppressive mediation, however, the case of effect unmasking (Case C in Figure 1, above), is easily confused with specification/moderation. The reason for the confusion may be that in some cases of moderation the overall bivariate X—Y relationship is zero. This is a particularly important case, because we might wrongly conclude that X and Y are unrelated, whereas they are related, but function differently at different levels of the moderator variable, Z. A similar mistake might be made when effect unmasking suppression occurs: a zero bivariate correlation for X and Y might lead a researcher to give up on assigning a potentially causal role to X, although
for different reasons than when Z moderates the X—Y effect. Extraneousness is a case of particular general importance in which the “true” X—Y relationship is zero. This paper focuses on the conceptual differences between extraneousness and various outcomes involving mediation.

One way to clarify the differences between instances of extraneousness, mediation, and moderation associated with a zero X—Y relationship is to examine the scatterplots and regression lines for each outcome. Figure 3 shows scatterplots and regression lines for hypothetical (but realistic) data that depict three different data outcomes: extraneousness (Case A), moderation by Z (Case B), and effect unmasking suppression, a case of mediation by Z (Case C).

**Extraneousness.** The regression line for extraneousness, Case A (top) in Figure 3, is based on all data points and reflects a positive X—Y relationship ($r = .52$). But the correlations at each level of Z, the antecedent variable, are very small and nonsignificant. The corresponding path diagram, shown to the right of the scatterplot, reveals that the path estimate for X is essentially zero ($-.003$), and that Z has strong positive correlations with both X and Y. The plot and diagram provide another regression-based perspective on the key outcome of extraneousness, in which X plays no causal role at all in predicting Y.

**Moderation by Z.** The middle example, Case B in Figure 3, shows a scatterplot representing an outcome of moderation for a bivariate X—Y relationship that is near zero and nonsignificant ($r = -.12$). It is important to note, however, that at the lowest value of the moderator, Z, the X—Y relationship is highly positive ($r = .75$), whereas at the highest value of Z, the X—Y correlation is
markedly negative \((r = -0.66)\). For the intermediate values of \(Z\), the \(X—Y\) correlation is small.

\(Z\) moderates the \(X—Y\) relationship in this example because the relationship of \(X\) and \(Y\) is different at different values of \(Z\): one time highly positive, one time highly negative, and other times, very small. The associated path diagram for these data (to the right of the scatterplot) shows that all the simple correlations are fairly small. In an idealized situation, all three correlations could be exactly zero and a similar moderation pattern would be apparent in the scatterplot for each level of \(Z\).

For these hypothetical data, a moderated regression analysis – the appropriate statistical test for detecting moderation (interaction) in such data – yielded a significant interaction product term \((p < .001)\) that increased the \(R^2\) squared by .20 over the nonsignificant 4% of variance explained by the combined (main) effects of \(X\) and \(Z\) alone.

Mediation by \(Z\). The bottom scatterplot, Case C in Figure 3, depicts effect unmasking suppression, one form of mediation by \(Z\) and a pattern sometimes confused with the preceding instance of moderation. In this case, we see that the overall \(X—Y\) relationship is again nearly zero \((r = 0.072)\). The relationships at each level of \(Z\), however, are similar and positive, ranging only from .14 to .37. An idealized, but less realistic, case could have been shown for which the positive \(X—Y\) correlations were identical at each level of \(Z\), but \(X—Y\) overall was zero.
In the case of suppression, the relationships at each level of Z are very similar in size and direction, in contrast to moderation, when they differ significantly from one another. Another distinguishing characteristic is that, in suppression, Z has a definite temporal relationship to X (i.e., it clearly follows X in time), whereas for moderation it does not. In addition, the relationships of Z to X and Y must be significant and substantial for mediation to occur.

The associated path diagram to the right of the scatterplot shows the expected suppressive pattern for these data: the betas exceed their corresponding r’s, and Z’s direct effect on Y is of opposite sign (negative) to its indirect effect through X (positive).

6 Some Examples Involving Realistic or Actual Data

Figure 4 provides three hypothetical data examples depicting extraneousness (Case A), redundant recast mediation (Case B), suppressive recast mediation (Case C), and one actual data example showing suppressive mediation (Case D). Once again, these diagrams show path coefficients on the lines with the simple correlations for X and Z preceding the slash mark.

**Extraneousness.** In Case A, physical punishment or spanking in early childhood correlated .24 with later anti-social aggressiveness in adolescence. Controlling for the child’s irritability, a basic characteristic of temperament, resulted in the initial 0.24 relationship dwindling to nonsignificance (beta = .008).

**Redundant recast mediation.** In Case B of Figure 4, controlling for the socio-economic status of new mothers (Z) results in a diminishment of the initial relationship between regular visits to an obstetrician (X) and the baby’s health
score at birth (Y). Despite the diminished effect size, however, the path coefficient remains significant. The proper interpretation of these data, which reflect redundant recast mediation, is that the mother’s SES improves the chances of the baby’s being healthy, both directly and indirectly, because higher SES mothers are more likely to make regular prenatal visits to the doctor.

A more elaborated future model might examine what other mediating factors, unmeasured in the model presented here, account for the finding that higher SES causes better infant health. In an actual study with this purpose, it would be important, of course, to note that the X—Y relationship had survived as causal (regular physician visits remain beneficial). But the researcher would be remiss, and a fuller explanation would be lost, if he or she ignored the causal role of socio-economic status.

Suppressive recast mediation: A hypothetical example and some actual data. Case C in Figure 4 is the behavioral/clinical version of the fire truck example mentioned earlier. In this example of suppressive recast mediation (with sign change), we see that a three-variable model reveals (a) the beneficial effect of therapy and (b) the indirect contribution of problem severity, which drives the client into more therapy sessions.

Case D in Figure 4 is based on our unpublished data on client derogation or depersonalization, a dimension of worker burnout measured in the Maslach burnout inventory (Maslach & Jackson, 1981). In this example of suppressive mediation, number of years of job experience mediates the effect of worker age on depersonalization. Age of worker is related to more depersonalization,
because older worker have logged more years in practice. But directly (i.e., for other unmeasured reasons), older workers were less likely to depersonalize their clients.

An important additional point that might be added for the suppression cases C and D is their obvious practical and theoretical importance. Researchers have sometimes wrongly believed that suppression was a peculiar statistical artifact or condition, devoid of practical research interest. Koeske (1998) has discussed this unfortunate misunderstanding in the context of research on the outcomes of different parenting styles.

_Larger systems: More than two predictors._ As might be expected, the analysis of larger systems, including the process of characterizing and distinguishing them, becomes more complicated when the number of variables exceeds three. In a larger path diagram with more predictors and potential mediators, however, extraneousness is apparent if a primary X’s direct effect and indirect effects are nonsignificant. Determining redundancy and suppression becomes more difficult, since both forces might operate within the same system. Yet a system may be identifiable as “predominantly redundant” or “predominantly suppressive.” And, paths that represent suppressive versus redundant influences can be identified, based on the rule of opposing (suppression) versus compatible (redundant) signs for direct and indirect effects.

7 _Identifying and Assessing Mediation: A Final Note_

One purpose of this paper is to clarify the importance of mediation in multivariate tests that begin with a bivariate correlation of special interest. In a
Majority of such cases, mediation is the likely outcome that needs to be described and interpreted. The importance of mediation effects raises the analytic and interpretative issue of how to formally determine that mediation is operative. How do we decide that an indirect effect is “real?” A partial answer to this question is provided by the following rule-of-thumb: if the coefficients along a path are significant and substantial, mediation can be claimed. “Substantial” may be defined as at least .20.

MacKinnon, Lockwood, Hoffman, West and Sheets (2002) recently reviewed three formal procedures for assessing mediation. One of these, proposed by Baron and Kenny (1986), would require significant coefficients for X—Zi and Zi—Y, as well as a reduction in the X—Y relationship to zero or nonsignificance. A second approach reviewed by MacKinnon et al. involves the development of statistical tests that assess the significance of the difference between a simple correlation and a partial coefficient. The third approach involves testing the significance of indirect effects that are obtained by multiplying the path coefficients along a path.

MacKinnon et al. (2002) considered some of the specific tests based on this third approach to be a good compromise of Type I error and power, compared to the other two approaches. The reader should consult MacKinnon et al. for guiding formulas to formally assess mediation. In lieu of their formal and technical approach, the above-stated rule-of-thumb may be a sufficient guide for many research situations.

**8 Summary and Conclusions**
We have attempted to address several issues we feel are important in the conceptualization and practice of research. These may be summarized as follows:

1. The teaching and use of common variable language – a comprehensive lexicon of variable types – is essential for both effective communication in research and for the production of quality research.

2. The classical variable typologies developed in sociology and psychology have historically used different terms for the same processes, but recently a more uniform set of terms has begun to evolve. We feel that it is incomplete, that a more inclusive synthesis is needed, and that students of research should be exposed to it.

3. We offer a variable typology in which extraneousness is but one of several possible patterns of three-variable relationship involving an antecedent variable, Z. We suggest that the disconfirmation of extraneousness, a frequent outcome in research, should motivate the exploration of alternative variable patterns, all of which involve mediation.

4. Simple multiple regression and path analytic terms are used to frame the synthesis and typology we present, since these terms best reflect the system properties of the variable networks discussed. Path diagrams and/or regression lines are used to clarify the features of each three-variable system described, and to distinguish three often confused outcomes: extraneousness, moderation by Z, and mediation by Z, in the case of an overall X—Y relationship of zero. Distinguishing these three possibilities is
important and their distinctiveness has not always been clear in the research literature readily accessible to students and practicing researchers.

5. X, the independent variable, retains a causal role in each of three possible variable systems involving an antecedent Z in which Z is not extraneous, but X must be recast as a mediator of the Z—Y relationship. Four additional patterns of relationship result when Z is instead understood as intervening between X and Y, rather than as antecedent to X. All of these latter patterns involve mediation or its disconfirmation.

6. Suppression can be understood as a special case of mediation in which the direct and indirect effects of a prior variable have opposite signs. Examples of suppressive systems that suggest the practical and theoretical importance of suppressive effects in many research scenarios are provided.

7. A recent article that evaluates different statistical procedures for quantitatively assessing the presence and magnitude of mediation is cited. A simple rule-of-thumb for identifying mediation is suggested for current practice: the path coefficients along a path must each be significant and substantial. Suppressive mediation is present whenever the sign of one path to Y is opposite to the sign of another path to Y (including a direct path).
Footnotes

1 The less frequent condition of a coterminous X and Z is omitted in our presentation. If X and Z were uncorrelated, there would be no shared explanation, and the situation would be uninteresting. If both predictors of Y were zero or non-significant, no causal effects could occur and, again, the situation would be uninteresting.
References


A comparison of methods to test mediation and other intervening variable effects.

*Psychological Methods, 7, 83-104.*


*Journal of Occupational Behavior, 2, 99-113.*


Figure Captions

*Figure 1.* Four causal patterns occurring when the third variable, $Z_a$, is an antecedent variable.

*Figure 2.* Four causal patterns occurring when the third variable, $Z_i$, is an intervening variable.

*Figure 3.* Scatterplots with regression lines showing extraneousness, moderation, and suppression.

*Figure 4.* Substantive hypothetical and real examples of extraneousness, redundant recast mediation, suppressive recast mediation and suppressive mediation.
Z is Prior to X, Za

A. Extraneousness

\[
\begin{array}{c}
\text{Z} \\
\text{Z} \\
\text{X} \\
\text{Y}
\end{array}
\]

B. Redundant Recast Mediation

\[
\begin{array}{c}
\text{Z} \\
\text{Z} \\
\text{X} \\
\text{Y}
\end{array}
\]

C. Suppressive Recast Mediation

\[
\begin{array}{c}
\text{Z} \\
\text{Z} \\
\text{X} \\
\text{Y}
\end{array}
\]

D. Simple Recast Mediation

\[
\begin{array}{c}
\text{Z} \\
\text{Z} \\
\text{X} \\
\text{Y}
\end{array}
\]

Z→Y correlation significant

Z→Y correlation was 0
Z is Consequent of X, Zi

E. Classical Mediation

F. Redundant Mediation

G. Suppressive Mediation

H. Disconfirmed Mediation

Z–Y correlation significant

Z–Y correlation was 0
Case A

Spuriousness Example

Not Significant in Each Level of Z

Moderation Example

Not Significant Over All Cases

Case B
Effect Unmasking Suppression
Not Significant For All Cases

+ Relationship For Each Z Level

Case C
A. Extraneousness

- Irritable Temperament, Za
  - Spanking/Physical Punishment
    - Adolescent Aggressiveness

B. Redundant Recast Mediation

- Socioeconomic Status, Za
  - Regular Visits to Physician
    - Baby's Health

C. Suppressive Recast Mediation

- Severity of Problem, Za
  - Number of Therapy Sessions
    - Mental Health Symptoms

D. Suppressive Mediation

- Age of Worker
  - Depersonalization of Clients
    - Years of Job Experience, ZI