

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600



A

**WHY DID THE CHICKEN CROSS THE ROAD?
EVALUATING THE VALIDITY OF CAUSAL INFERENCES**

by

ELLIE A. BUTEAU

**A dissertation submitted to the Graduate Faculty in Psychology
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy, The City University of New York**

2003

UMI Number: 3083646

Copyright 2003 by
Buteau, Ellie A.

All rights reserved.

UMI[®]

UMI Microform 3083646

Copyright 2003 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2003

ELLIE A. BUTEAU

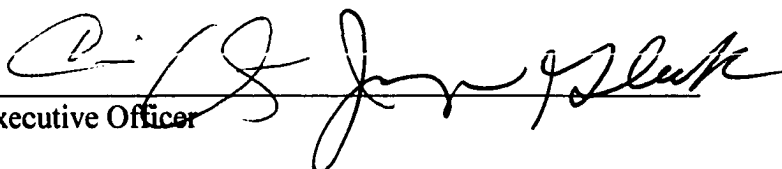
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

4/15/03
Date

David Rindskopf
Chair of Examining Committee

4/15/03
Date


Executive Officer

David Rindskopf

Michelle Fine

Vita Rabinowitz

Keith Markus

Martin Ruck

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

WHY DID THE CHICKEN CROSS THE ROAD? EVALUATING THE VALIDITY OF CAUSAL INFERENCES

by

Ellie A. Buteau

Advisor: Professor David Rindskopf

The concept of causality remains a controversial issue in social scientific research. For decades, researchers have used the Elaboration Model, established in sociology, and the Validity Typology, established in psychology, as frameworks for analyzing causal inferences. Rubin's Causal Model, developed in statistics and not widely known in the social sciences, offers a system of causal effect equations and a causal variable definition that contribute interesting insights to the evaluation of causal inferences. Each of these three models was applied to three published research articles. The comparison of applications demonstrates some disagreement over what constitutes a valid causal inference, and also highlights areas in each model that could benefit from clarification. A resulting list of guidelines was developed for researchers to consider when designing and interpreting research meant for causal inferences.

Acknowledgements

I would like to thank my advisor and mentor, Professor David Rindskopf, for his continued support, guidance, encouragement and patience throughout my work on this dissertation. His insight about my interest in causality and social variables resulted from his caring enough to listen to my ideas; he helped me translate those ideas into a project I enjoyed pursuing for almost three years. The development of this dissertation resulted from our countless conversations and idea exchanges during which I developed important problem solving skills necessary to complete this work. Without his dedication to seeing this project through and his commitment to my development as a scholar, I would not have reached where I am today. His philosophies about the research process and about life itself helped me put my work into perspective, and taught me to appreciate the complexities and challenges of my research.

In addition, I thank Professor Vita Rabinowitz for her comments on earlier drafts of this work and her interest in issues of methodology. Her ideas led to more thorough explanations of methodological concepts sometimes taken for granted and to my desire to clarify the purpose and applications of my research. Her enthusiasm in the development of this work contributed to its evolution.

I also thank Professor Michelle Fine for her encouragement to more deeply pursue the conceptual richness of this topic. Her invitation to discuss this work in her classes resulted in my better understanding my positioning on topics in this paper. The formation of my thoughts about race, ethnicity, and sex in this research resulted in large part from her influence on my intellectual development. Professor Fine's belief in this work promoted my confidence in its inherent political stance.

With much respect and sadness, I thank the late Sue Rosenberg Zalk for her teachings, and her inspiration to contribute to the study of gender, race and issues that shape our everyday lives. Her bold spirit and unwavering commitment to work for the causes she believed in influenced the path that led me to this project. Her comments on earlier drafts of this work sustained the purpose I set out to pursue.

I also thank Professor Keith Markus for his involvement in my dissertation committee, and his helpful comments and questions about my work. His attention to assumptions inherent in my thinking encouraged me to be more precise in my own language choices. I thank Professor Martin Ruck for his comments on a draft of this paper and his involvement in my dissertation committee as well.

With immeasurable gratitude, I thank my Mom and Dad for instilling the value of education in me and for believing in the educational goals I set out to pursue. Their confidence in my abilities and their belief in the worthiness of my interests carried me through many difficult transitions. Their multiple paths of support, since the beginning of my education and continuing through college and graduate school, enabled me to keep progressing and succeeding, and their pride in my accomplishments kept me focused when I felt my motivation waning.

And last, but certainly not least, I thank Buddy for his silent yet imperative support; his companionship sustained me through the many hours upon hours of work comprising this project.

Table of Contents

Title Page		i
Copyright Page		ii
Approval Page		iii
Abstract		iv
Acknowledgements		v
Table of Contents		vii
List of Tables		xi
Chapter 1: Introduction		1
Criteria for Causal Relationship		1
Validity		2
Methodological Models		3
Statement of Question		6
Organization of Dissertation		6
Chapter 2: Methods of Control		8
Increasing Precision and Reducing Bias		9
Research Design		10
Statistical Control		14
Chapter 3: Rubin's Causal Model		17
Causal Effect and Causal Variable		18
Study Types		21
Process of Analyzing Causal Effects		22
Criticisms of Rubin's Causal Model		39

Chapter 4: Elaboration Model		62
Test Variables and Partial Relationships		63
Chapter 5: Validity Typology		76
Types of Causal Conclusions		77
Causal Variables		78
Four Types of Validity		79
Notation		82
Chapter 6: Regression Framework		85
Choosing Control Variables		88
Statistical Models		90
Structural Equation Modeling		94
Relationship with Other Models in this Paper		95
Chapter 7: Comparing and Contrasting RCM, EM and VT		97
Defining a Causal Variable		97
Types of Studies and Relations		98
Differing Foci		98
Assumptions versus Threats		99
Acknowledging the Role of Theory		100
Involving Mathematics		101
Representation of Data		102
Application to Various Research Designs		102
Chapter 8: Language		104
Creating Scientific Knowledge		105

Interpreting Language	107
Chapter 9: Method	110
Data	110
Procedure	114
Interpreting the Data	124
Chapter 10: Analysis of Article 1: Hill (2001), “Parenting and Academic Socialization as they Relate to School Readiness: The Roles of Ethnicity and Family Income”	125
Causal Language	125
Applying Rubin’s Causal Model	128
Applying the Elaboration Model	142
Applying the Validity Typology	147
Applying the Regression Framework	152
Summary of Results after Applying Each Model to Article 1	154
Chapter 11: Analysis of Article 2: Wayne, Riordan & Thomas (2001), “Is All Sexual Harassment the same? Mock Juror Decisions in Same- and Cross-Gender Cases”	156
Causal Language	157
Applying Rubin’s Causal Model	158
Applying the Elaboration Model	167
Applying the Validity Typology	172
Applying the Regression Framework	175
Summary of Results after Applying Each Model to Article 2	176

Chapter 12: Analysis of Article 3: Anderson, John, Keltner & Kring (2001),		
“Who Attains Social Status? Effects of Personality and Physical		
Attractiveness in Social Groups”		179
Causal Language		180
Applying Rubin’s Causal Model		182
Applying the Elaboration Model		189
Applying the Validity Typology		193
Applying the Regression Framework		196
Summary of Results after Applying Each Model to Article 3		198
Chapter 13: Discussion		200
Language		202
Elaboration Model		203
Validity Typology		207
Rubin’s Causal Model		211
Limitations		235
Guidelines		236
Chapter 14: Conclusion		241
Avoiding Misinterpretation and Misuse		241
Summary		245
Tables		246
References		260

List of Tables

Table 1: Notation for Rubin's Causal Model	246
Table 2: Example of a Randomized Study	248
Table 3: Example of a Randomized Study Using Rubin's Causal Model	249
Table 4: Example of a Nonrandomized Study Using Rubin's Causal Model Notation	250
Table 5: Frequencies for Bivariate Relationship between Religion and Age	251
Table 6: Partial Relationships Formed from the Test Variable 'Sex'	252
Table 7: Partial Relationships Classification for the Elaboration Model	253
Table 8: Examples of Causal Language	254
Table 9: Examples of Associational, Predictive and Ambiguous Language	255
Table 10: Bivariate Relationship between Harasser and Verdict	256
Table 11: The Role of Gender Combination in Relation to Harasser Gender and Verdict	257
Table 12: Summary of Comparisons among the Three Models After Applications to Articles	258

CHAPTER 1: INTRODUCTION

Questions studied by today's social scientists continue to grow more complex and profound. From examining racial discrimination in education, to achieving peace between countries, explanations offered by researchers hold the power to contribute insights necessary for constructive change. These explanations result from an intricate process of analyzing potential causes, evaluating effects, and ultimately drawing valid causal inferences.

The goal of this dissertation is to analyze the logic and methods social scientists currently use to form causal inferences, particularly in situations involving complex variables such as race, sex, and personality. Causal inferences shape future research and, to some extent, influence policy formation (e.g., Fine & Barreras, 2001; Rein & Winship, 1999; Uggen & Piliavin, 1998) and inform testimony in courts of law (e.g., Fiske, Bersoff, Borgida, Deaux & Heilman, 1991; Jasanoff, 2001; Zeisel & Kaye, 1997). Such applications of these inferences require careful consideration of how we define causal relationships and assess the validity of causal conclusions. Contemplating issues of causation continues to be an essential area of advancement in social science research (Shadish, 1995).

Criteria for Causal Relationship

Philosophers and scientists have been contemplating the issue of causation for thousands of years (Note: See Holland, 1986a and Cook & Campbell, 1979, for historical reviews of relevant philosophical discussions about cause). What exactly is a researcher searching for when trying to identify a cause? How is a causal relationship defined? In the literature, researchers discuss certain conditions that must be met in order to consider

labeling a relationship as causal (e.g., Blalock, 1961a; Cook & Campbell, 1979; Cox, 1992; Holland, 1986a; 1986b).

Covariation comprises a necessary aspect of a causal relationship: Two variables must covary in order to be considered related in a causal way. Temporal order, or the direction of the relationship, defines another necessary aspect of a cause-effect relationship. An effect must occur after its cause has taken place. In research, we cannot always observe this temporal order and must sometimes turn to theory to help specify relationships among variables (Cox, 1992). Once a causal relationship has been identified, it must continue to exist in the presence of a third variable. A cause-effect relationship that disappears upon the introduction of a third variable is labeled as a spurious relationship. These criteria have been discussed by philosophers for centuries and their fundamental role in establishing causal relations is agreed upon across disciplines, as well as by proponents of both quantitative and qualitative research methods (Shadish, 1995).

Validity

Although these causal criteria seem basic, consistency in satisfying them becomes more complicated when studying people's behaviors, characteristics, and thoughts, as opposed to objects and conditions in the physical sciences, or even animal behavior to some extent. People do not act and react without thinking, feeling, and calling on memories and past experiences. Influences from contexts and environments also contribute to behaviors and attitudes. As world events, science and technology evolve quickly, so do our attitudes and habits. But, even with these myriad causes acting in conjunction, well-designed theories and methods can guide us in specifying causes.

Social scientists build knowledge about cause-effect relationships based on confidence in the validity of empirical evidence (Marini & Singer, 1988).

What is meant by the term validity? There are various types of validity. Validity has commonly been discussed from two aspects: measurement (i.e., psychometrics) and inferences (Ghiselli, Campbell & Zedeck, 1981). This dissertation concerns the validity of inferences. Validity refers to truthfulness and accurateness, and an inference is a conclusion deduced from empirical evidence. The validity of inferences concerns the degree of truth and accuracy of conclusions (Shadish, Cook & Campbell, 2002).

Methodological Models

What models exist for a methodological analysis of the validity of causal inferences? Various models differ in their approaches, sometimes leading to disagreement on what constitutes a valid causal statement. These models focus on differing aspects of the research and data analytic processes.

This paper focuses on three methodological approaches to evaluating causal inferences, used to varying degrees in current research methodology classes and research practice. Two of these models, the Validity Typology and the Elaboration Model, are established approaches to causal analysis in the social sciences, while the other model, Rubin's Causal Model, is fairly new to social science research. The reason for examining Rubin's Model concerns the discrepancy between some of its stances and common social scientific practices.

In psychology, an established approach is the Validity Typology (VT). Campbell first developed this approach in the late 1950s. Early foundations of the model appeared in the book "Experimental and quasi-experimental designs for research" by Campbell &

Stanley (1966). That book explained possible threats to validity, logistical as well as theory-driven, that a researcher must consider when designing or analyzing a study of causal relationships. Updated and enhanced explanations of this approach continued to be published, the most recent being Shadish, Cook & Campbell (2002).

In sociology, a main approach is the Elaboration Model paradigm (EM), originally developed by Lazarsfeld and Kendall after WWII. Elaboration takes place after a relationship has been established between two variables and the researcher wishes to examine why and under what conditions that relationship exists. Babbie (e.g., 2001) explains the basic tenet of this model as clarifying causal effects through multivariate relationships.

A third approach to be explored in this paper is Rubin's Causal Model (RCM). Beginning in the 1970s, Donald Rubin, a statistician, developed this conceptual model, outlining criteria for analyzing causal effects. Continuing through the 70s, 80s and 90s, Rubin, as well as other statisticians and econometricians applied and extended this model.

Rubin's Causal Model presents a framework of definitions and equations developed to assess the validity of causal inferences (Holland, 1986a; 1986b; Holland & Rubin, 1983; Rubin, 1974). RCM does not provide a statistical method for analyzing data; it is a framework for explicating assumptions implicit in causal interpretations of data. Using this model to evaluate the possibilities of testing implicit assumptions in a study, the path to arriving at causal conclusions can either be validated or invalidated.

Of key importance to this dissertation is Rubin's Model's definition of a causal variable. It does not classify variables such as 'sex,' 'race,' and 'age' as causal variables, and this presents critical challenges to the work of social scientists. A large portion of this

dissertation is dedicated to better understanding the causal variable issue and providing researchers with alternative ways to conceive of this problem.

In more recent years, Rubin's Causal Model received more attention outside of the statistical literature. Sociologists showed increased interest in this model for its use in analyzing causal effects in observational data (e.g., Marini & Singer, 1988; Uggen & Piliavin, 1998; Winship & Morgan, 1999). In the latest edition of the Handbook of Research Methods in Social and Personality Psychology, West, Biesanz, & Pitts (2000) offer a description of the model in less technical terms. Most recently, Shadish (in press) stated the need for more critical work on Rubin's Causal Model, stressing that field experimentation researchers ought to know this model well.

While the VT and EM paradigms are widely used to assess causal inference in social science work, to this point RCM has not received much in-depth attention in the social science disciplines. This may be because of the technical nature of the model (e.g., equations). Some have criticized Rubin's Causal Model for not contributing any new insight into causal analysis. Leamer (1988), for example, explicitly questions whether or not researchers should take the time to learn this model because its utility above and beyond what researchers already know about causal analysis is not apparent. (The nature of these criticisms will be discussed in detail later.) In contrast, Shadish (in press) considers this model one of the most important advancements in observational study research in recent decades. After an initial read, the contribution of the model *can* easily be missed because its insights build upon basic tenets of research methods and design issues such as randomization.

This dissertation concentrates on Rubin's Causal Model and what, if anything, it offers an analysis of causal inferences above and beyond the already accepted Validity Typology and Elaboration Model. Each model approaches the issue of causality with its own unique emphases. What can social scientists learn from the approach put forth by Rubin's Model?

Statement of Question

What methodological considerations does Rubin's Causal Model contribute to social science research? To answer this question, RCM will be compared with the two most widely accepted and utilized methodological models in social sciences, the Validity Typology and Elaboration Model to explore:

- a) What does each model offer a researcher who wishes to judge the truthfulness/accuracy of a causal inference in a study? Specifically, what guidance does each model offer to a researcher drawing causal inferences about demographic (e.g., sex, race) and personality variables?
- b) What elements of the causal analysis process does each deem important/emphasize when judging the validity of causal inferences?
- c) Are there situations in which one model would be more useful than the others for an analysis of the validity of causal inferences?

Organization of Dissertation

To answer these questions, this paper will be organized in the following way: When reaching causal conclusions, ruling out third-variable explanations becomes critical; therefore, methods for ruling out third-variable explanations will first be discussed. Then, each of these three models will be separately defined and explained. A

fourth model, the regression framework, will also be discussed because it is a common choice for examining data to answer causal questions regardless of research design. Then, the VT, EM, and RCM will be compared and contrasted to identify their unique and common features.

Next, steps will be outlined for an analysis of published research through the implementation of each model. RCM, VT, EM and a fourth model, the Regression Framework, will be applied to published research studies to examine what unique and/or overlapping information each model contributes to the study of valid causal inference. In the discussion chapter, I present a list of guidelines for researchers to consider when designing and interpreting studies meant for casual inference. In the conclusion chapter, I summarize the contribution of each model, as well as the contribution of concepts introduced in this dissertation, to planning and evaluating a study meant for causal inference.

CHAPTER 2: METHODS OF CONTROL

As mentioned in the introduction, three main criteria exist for identifying a possible cause-effect relationship: covariation, temporal order, and ruling out third variable explanations. Ruling out possible third variable hypotheses remains the most difficult criterion to achieve. In traditional approaches, only after the influence of additional variables has been considered can valid statements of causal conclusions and inference follow.

To accurately identify the causes of effects, researchers try to separate, or capture the influence of, relevant variables not under study to clearly examine the effects of the hypothesized causal variable(s). Variables that a researcher attempts to measure or separate for this purpose are known as ‘confounds’ (e.g., Cochran, 1983) or ‘control variables.’ The process of separating these variables from the main variables of interest in a study is known as ‘controlling for’ a variable (Cook & Campbell, 1979). The foremost purpose of controlling for a variable is to rule out alternative hypotheses or eliminate “threats to valid inference” (Cook & Campbell, 1979).

A researcher must think ahead about which variables may exert an influence in a study (Cochran, 1983) and approach either the study design or analyses in a way that captures those influences. Researchers use two main methods to control for the effects of possibly confounding variables: research design and/or statistical control. These methods reduce the bias (i.e., inaccuracy) and increase the precision with which causal inferences are made.

[Note: The word ‘bias’ communicates the idea of a lack of objectivity or distorted perception. The type of bias controlled for by research design and statistical analysis is

only one type of bias in research, and attention to it should not imbue the results of a study with a false sense of impartiality. Other aspects of a study, such as which levels of treatment are chosen, the type of participants and outcomes selected, and which results get reported (e.g., trends, competing results) can also be considered as biasing the results of a study (Devault, 1996; Sherif, 1994; Smith, 1994)].

Increasing Precision and Reducing Bias

In this paper, bias and precision refer to two characteristics of a sample statistic (i.e., an estimate of a population parameter) about which one draws causal inferences. The statistic of interest in this paper is the estimate of the causal effect.

Precision

An estimate of a causal effect is precise if with repeated sampling from a population, the resulting estimates cluster closely together (Moore, 1997). If, with repeated samples, the resulting estimates vary considerably, demonstrating inconsistency in estimation, then these are not precise estimates. The idea of precision does not concern the accuracy of the estimate. The sample estimates can be very far from the population value, but as long as they are clustered together, demonstrating little variance in their estimations, they are precise estimates.

To consider what precision might look like, imagine a bulls-eye (citation) in the center of a circle. The bulls-eye represents the causal effect value of a variable in the population. Now imagine many dots around the bulls-eye, each representing the causal effect value estimated for one sample of participants on this same variable. If these dots cluster together, regardless of whether they are near or far from the bulls-eye, they are

precise estimates. If these dots are scattered around, and do not cluster together, the sample causal estimates are not precise.

Bias

Bias refers to the degree to which the sample estimate systematically differs from the population value. An estimate is biased if in repeated samples from a population, the sample estimates systematically differ from the population value. Biased estimates lack accuracy; unbiased estimates are accurate estimates. Referring to the example of the bulls-eye, the farther a dot appears from the bulls-eye (recall that the bulls-eye represents the population value of the causal effect estimate), the more biased this sample estimate is; dots closer to the bulls-eye represent more accurate estimates.

Precision refers to the location of a sample estimate (a dot) in relation to the other sample estimates; bias refers to the proximity of a sample estimate to the population value (the bulls-eye). An estimate can be simultaneously precise and biased, or simultaneously imprecise and not biased. Researchers can minimize bias and maximize precision by controlling for variables through research design and/or statistical analyses (Cochran, 1965, 1971; Cochran & Rubin, 1973).

Research Design

Researchers achieve control via research design using a number of techniques: How to select participants (e.g., from college classrooms or community centers), which participants to select (e.g., only those of a certain age or race), how variables are measured (e.g., self reports or observational), and the process of assigning participants to conditions (e.g., random assignment). Methods of assigning participants, to a certain degree, assure that many potentially biasing variables will be dispersed among

conditions, without influencing the outcomes of participants in one specific condition more than any other condition. The purpose of these design methods is to control for variance in the dependent variable contributed by variables other than the independent variable(s) of interest in the study.

The most effective strategy for distributing the influence of confounding variables among study conditions is random assignment of participants to conditions (Campbell & Stanley, 1966; Cook & Campbell, 1983). Random assignment refers to the process of placing participants into conditions when placement is unpredictable and not dependent upon any characteristics of the participants.

Random assignment is a methodological tool, implemented before collecting outcome data for a study. A psychologist running an experiment in a laboratory or in the field controls certain variables by randomly assigning participants to conditions. When possible, the influence of confounding variables should be minimized through designing a study that includes random assignment of research participants to groups (Cochran & Rubin, 1973).

Other methodological designs used to control for biasing variables include randomized block and matched-pair designs. Randomized blocks involves pairing participants on the basis of a characteristic(s) related to the outcome, and then randomly assigning one member of each pair to the treatment condition and the other member to the control condition. The pairs may be initially matched on demographic characteristics such as sex or race, or ability tests, such as IQ. This design offers more statistical power than simple random assignment without matching (Glass & Hopkins, 1996), because any

test of an outcome measurement will be more precise after equating groups on relevant confounding variables before collecting response measures.

In a matched-pairs design, participants are matched on the basis of characteristics related to the outcome variable, but unlike in the randomized block design, members of the pairs are not randomly assigned to the treatment and control conditions. The researcher matches the pairs after the treatment and control groups have already been formed. The matched-pairs design is usually used in a retrospective manner. For example, after the responses of members of the treatment condition have been measured, a control group may be formed to match the treatment group on the basis of characteristics relevant to the outcome variable.

Matched-pairs designs often lead to erroneous conclusions (Judd, Smith & Kidder, 1991). Differences in the outcome may exist because groups may not have been matched on all necessary variables. In addition, regression to the mean may result in conclusions that do not accurately represent the treatment effect (Glass & Hopkins, 1996; Judd et al., 1991). If groups were matched on the basis of a single pretest score, regression to the mean could lead to incorrect conclusions about the effect of the treatment on the posttest scores; scores on the posttest may be explained by regression to the mean rather than an effect of treatment.

Random assignment takes care of most bias problems (Cochran, 1965; Cochran & Rubin, 1973; Cook & Campbell, 1979; Moore, 1997), allowing the researcher to conclude that differences between the study groups on the response are due to the independent variable of interest. But room for increased precision may still exist in randomized studies (Cochran, 1965). Even when using random assignment to reduce the

likelihood of differences between groups, statistically adjusting for control variables can further improve the power of statistical tests (Boruch & Rindskopf, 1977). When deciding which, if any, variables to control for in a randomized study, Cochran (1971) suggests identifying only those variables that correlate at least .3 or .4 (Cochran, 1983) with the outcome. In the context of a randomized study, removing the variance of a variable that relates to the outcome increases precision of the estimation by reducing the error variance (Cochran, 1971).

For logistical and ethical reasons, random assignment and other techniques for assigning participants sometimes cannot be used, thus necessitating the use of quasi-experimental studies and passive-observational studies (Cook & Campbell, 1979) such as surveys. In nonrandomized studies, the effects of variables not controlled for remains a major concern (Boruch & Rindskopf, 1977; Cook & Campbell, 1979; Rubin, 1974). These studies often include groups formed by circumstances beyond the researcher's control. For example, the groups may be naturally occurring groups such as biological sex, or children of different ages, or groups existing prior to the research study (in which case matching would be a possible research design) such as classrooms of students, or self-selected groups (Cochran, 1983) such as political parties.

The primary concern with the use of such groups is that structure and membership of the groups cannot be controlled or changed by the researcher. In these contexts, a wealth of confounds usually exist (Cochran, 1965, 1983; Cochran & Rubin, 1973; Cohen & Cohen, 1983); variables related to group formation, group characteristics, shared beliefs, and inter-group relations, complicate the process of concluding that any effects of a treatment may be due to the independent variable of interest.

In quasi-experimental and nonexperimental studies, reducing bias becomes a primary concern for drawing accurate conclusions (Cochran, 1983). Bias exists in the resulting causal effect estimate when a study includes groups with such preexisting differences. The estimate would not accurately portray how the treatment variable affected the participants in conditions because confounding variables could be at least partly responsible for differences in responses. In studies of such groups for which randomization is impossible, statistical control becomes an indispensable tool.

Statistical Control

When variables cannot be controlled through research design, and sometimes even when they can be, researchers rely on the idea of statistically controlling for variables. A *control variable* (covariate) is a type of independent variable or predictor variable. Control variables differ from the independent variables of interest in a study in that their effects are not the focus of the study.

Definition

Statistical control is a broad concept. The main idea refers to the process of using statistical analyses to capture and isolate the variance in responses due to particular independent variables (Cohen & Cohen, 1983). Using statistical control, a researcher eliminates variance in the outcome not due to the independent variables of interest. The goal of statistically adjusting for control variables is to ensure that a variable other than the independent variable of interest is not responsible for the effects (Cochran, 1983), thus building confidence in the validity of causal conclusions.

Even after controlling for measured variables, treatment effects may still be biased by systematic differences due to additional unmeasured variables (Rosenbaum,

1984a). This bias arises when using statistical control as well as when using research design to achieve control. For example, when a researcher uses a matched-pair design, other variables not selected for matching may bias the results (e.g., Judd et al., 1991). These remaining non-matched biasing variables may, even after matching on some other variables, lead to biased results. Even implementing statistical control in the analysis for the matched-pair design may not eliminate the bias. It is crucial to remember that a data analytic tool such as statistically controlling for a variable cannot always be used to fix problems inherent in design (Boruch & Rindskopf, 1977).

Interpreting statistical control

When a researcher reports that a variable in a study has been statistically controlled for, what does that really mean? How would one interpret the statement, “...after controlling for family income, parental education, age and gender, satisfaction with social support was not related to locus of control” (Zea, Jarama, & Bianchi, 1995, p. 523). In a broad sense, controlling for ‘variable x’ means that the variance of ‘variable x’ has been statistically separated from the study’s main variables of interest. If the control variable relates highly to the outcome, removing its variance increases the precision of the resulting causal effect estimate, [i.e., decreases the amount of error in the statistical analyses and allows for a clearer picture of the effects of the independent variable of interest on the dependent variable (Cochran, 1971)].

Although the goal of controlling for variables is to arrive at more valid causal conclusions, instances of the inappropriate use of statistical control can lead to misleading or distorted outcomes (Boruch & Rindskopf, 1977; Cohen & Cohen, 1983; Evans & Anastasio, 1968; Holland & Rubin, 1983; Kahneman, 1965), less precise causal

estimates (Cochran, 1983; Cohen & Cohen, 1983), and conclusions that are difficult to interpret (Holland & Rubin, 1983; Lord, 1969). Therefore, a researcher needs a well thought-out model underlying the choice to control for a particular variable (Blalock, 1961b; Rubin, 1974).

These methods of controlling for third variable explanations via research design and statistical analysis provide the conceptual bases for the models of causal inference to be described in this paper. The models themselves supply systems of ideas and steps that, when followed, lead to increased confidence in the accuracy of a causal inference. Next, starting with Rubin's Causal Model, each methodological model of interest will be individually explained.

CHAPTER 3: RUBIN'S CAUSAL MODEL

Both the Validity Typology and the Elaboration Model mainly deal with the causes of effects. Both of these models concentrate on considering or ruling out potential alternative explanations (causes) for an outcome. Holland clearly states that the interest of Rubin's Model resides in the effects of causes, not the causes of effects (e.g., Holland, 1986a, 1986b). That is, this model focuses on measuring participants' *responses* in a study: How participants would potentially respond after being exposed to the possible causes of study.

For example, if a researcher wishes to determine the attitudinal effect of a violence prevention program, she might develop a study with two conditions, a treatment condition in which children are exposed to the program, and a control condition in which children are not exposed to the program. Exposure to the program is a variable manipulated by the researcher.

Rubin's Model suggests the question of whether or not the program works would be answered by comparing a child's attitude about violence after participating in the treatment condition to the child's attitude had she participated in the control condition. The difference in attitude after exposure to two conditions would equal the causal effect of the program for an individual child.

As exemplified through this scenario, in order to understand the causal effect of this violence prevention program, we must be able to consider a child's response to both conditions of the study in order to calculate the program's causal effect on the child. If it can only be imagined that the child could be exposed to one level of the program then

there would be no comparison information available to assess a causal effect. A main tenet of Rubin's Model dictates that a person must be *potentially exposable* to more than a single value of a variable if that variable is to be considered causal. The importance of the word 'potentially' indicates the recognition that in most studies each participant can only be exposed to a single condition. But one could imagine that it *would have been possible* for participants to have participated in a different condition of the study.

In order to imagine that a person's responses could be compared between conditions, the level of the variable the person is responding to must be one that can be changed, or manipulated (Holland, 1986a, 1986b; Holland & Rubin, 1983; Rosenbaum, 1984a; Rubin, 1974). RCM bases the process of determining causal effects on the concept of experimental manipulation (Holland 1986a; Holland & Rubin, 1983). When a researcher experimentally manipulates a variable, the researcher controls whether or not, and at what level, a participant will be exposed to this variable (Judd et al., 1991).

Causal Effect and Causal Variable

The comparison of how a participant would respond to one condition relative to another condition creates the basis for the model's definition of a causal *effect* and its definition for a causal *variable*¹. If the value of a variable to which a person is exposed could not be changed, if a person could not potentially be exposed to another value, Rubin's Model does not consider this variable a causal variable. Only variables for which a person can be exposed to more than one value can be causal because, as the analysis of causal effects dictates, the process of analyzing causal effects requires that the potential

¹ Neyman (1923) has been cited as being the original person to have used the concept of potential responses for comparing among conditions.

responses from more than one condition be compared. This comparison of responses takes place at the individual level, in theory, and at the group level in practice (this idea will be more fully developed later).

Holland & Rubin (1983) define a causal effect for an individual participating in a study with only two conditions as follows: The causal effect of treatment for an individual in a study would be derived by comparing what this person's response would be after exposure to the treatment condition, to what this person's response would be after exposure to the control condition. The difference between this individual's responses under the two conditions would be attributed to the presence versus absence of the treatment and defined as the causal effect of the treatment.

Following this definition of a causal effect, the definition of a causal variable dictates that a variable only be classified as causal if a person can potentially be exposed to all relevant levels of the variable. For example, the variable 'stereotype threat' (Steele, 1997) qualifies as a causal variable because a participant could be potentially exposed to, and have an outcome for, both the 'being threatened with the stereotype' condition, and the 'not being threatened with the stereotype' condition in a study. The variable 'biological sex' does not qualify as a causal variable because its two conditions, man and woman, are not both ones to which each participant could be potentially exposed. A man could not be potentially exposed to, and have an outcome for both the 'being a man' condition and the 'being a woman' condition in a study; he could only be exposed to the 'man' condition.

Attribute

A variable such as ‘sex,’ as well as variables such as ‘race’ and ‘age,’ are labeled as *attributes* in Rubin’s Model (Holland, 1986a; Holland & Rubin, 1983). These variables have also been referred to as intrinsic properties (Cox, 1992). Attributes are not considered causal variables in the framework of RCM. Although Rubin and Holland agree on this idea, they do not reach complete consensus on their reasoning. Holland (1986a; 1986b; 1986c) purports that attributes cannot be causal because a person could not potentially be exposed to a value of such a variable other than the value s/he currently possesses without other characteristics of her/him changing in the process of taking on this other value. Rubin (1986) more specifically explains why ‘male’ and ‘female’ cannot be causal. Without referring to the actual variable (sex), he provides the following scenario and explanation:

“If the females at firm *f* had been male, their starting salaries would have averaged 20% higher” (Rubin, 1986, p. 962). He states that Holland would not support this as a causal statement because it involves an attribute (i.e., sex). Rubin himself agrees that this does not classify as a causally meaningful statement, but not simply because it involves an attribute.

The problem for Rubin pertains to how to define the ‘male’ treatment. To solve the causal problem, he suggests a study, “considering the units to be the female employees at entry and the treatments to be ‘female,’ which is well defined since the units are females, and ‘male,’ which has many possible versions ranging from some hypothetical ‘at conception X to Y chromosome treatment’ to replacing ‘F’ with an ‘M’ on a job application form. Certainly these different versions of the treatment ‘male’ could

lead to vastly different outcomes” because the researcher would have to understand “which version of the treatment ‘maleness’ is under study or agreement on a way to average over some collection of such versions” (p. 962).

Upon reading this example, one immediately wonders how Rubin defines the treatment ‘female’ when he labels it as ‘well-defined.’ Rubin proceeds to offer another study to address the issue presented in his above example. He turns attention to studying hiring practices and takes attention away from the gender component altogether. He later refers the reader to an earlier article in which he states, like Holland, “Unalterable characteristics of individual units, however, such as age or sex, are not treatments: age and sex do not have causal effects because it is not meaningful to discuss the consequences of altering these unalterable characteristics” (Rosenbaum & Rubin, 1984, p.26).

This distinction of causal vs. non-causal variables does not lead to simple decisions as to what should be categorized as a cause, versus what should be categorized as an attribute. Holland (1986a) attempts to clarify this decision process by writing, “The general problem, I think, is in deciding when something is an *attribute* of units and when it is a *cause* that can act on units” (p.955, italics part of original quote).

Study Types

Using the inclusion of a causal variable as the indicator for the types of inferences drawn from a study, Holland & Rubin (1983) classify three types of studies: descriptive studies, uncontrolled causal studies, and controlled causal studies. “Descriptive questions differ from causal questions in that there is no implied comparison of the values of an

outcome variable under different levels of an experimental manipulation” (Holland & Rubin, 1983, p. 19).

The process of arriving at conclusions through their network of causal effect equations (to be described below) remains the same for descriptive questions, but must involve testable or untestable assumptions (Holland & Rubin, 1983). When a study involves attributes rather than causal variables as independent variables, associational, as opposed to causal, relationships are studied (Holland, 1986a) and descriptive, as opposed to causal inferences must be drawn (Holland & Rubin, 1983).

An uncontrolled causal study includes a causal variable not under the researcher’s control. Choice of town to live in or college to attend, or team membership would qualify as causal variables not under the researcher’s control. Causal inferences can be drawn from uncontrolled causal studies.

Controlled causal studies include a causal variable under the researcher’s control. The example from above, of a researcher exposing or not exposing participants to stereotype threat, qualifies as a controlled causal study.

Process of Analyzing Causal Effects

Observing a participant’s performance under varying conditions allows a researcher to understand how exposure to conditions differentially influences an outcome. As discussed above, “the experimental manipulation” is the focus of causal inferences in RCM (Holland & Rubin, 1983, p. 5). The definition of a causal effect requires that a person be potentially exposable to all values of a variable applied in a study because a causal effect is produced by comparing among a person’s potential responses to each value of the independent variable of interest.

This way of defining a causal effect applies at the individual level, and is, therefore, based on a criterion that does not occur in practice. A person can only participate in a single condition of a study, and a response can only be measured under one condition for every participant². Returning to the example above, if a child participated in a study evaluating a violence prevention program and this child was in the group receiving the treatment, it would be impossible for this child to fully experience the control condition for this study if the study was to be replicated; the child would already have the knowledge and perspective from having participated in the treatment group. Therefore, data for a participant is only available for one condition, presenting what can be thought of as a missing data problem.

The impossibility of analyzing causal effects at the individual level has been discussed as a problem in the context of causal inference, and labeled as the “Fundamental Problem of Causal Inference” (Holland, 1986a; Holland & Rubin, 1983). Given that we cannot usually observe an individual’s outcomes from more than one condition in a study, what does this mean for inferring causal effects in Rubin’s Causal Model?

Due to this fundamental problem, the *average* causal effect, not the *individual* causal effect, becomes the effect of interest. Whereas the individual causal effect results from comparing the responses of a single person, the average causal effect averages responses from all participants in a condition before comparing responses between conditions. The average causal effect is calculated using, for example, the average

² In certain designs, participants are exposed to all conditions (e.g., repeated measures designs). There can, however, be carry-over effects between conditions. In many studies, it is not possible to avoid carry-over effects, and participants can, therefore, only be measured under a single condition.

treatment condition response and the average control condition response (i.e., averaging over the individual response values in the treatment condition, and averaging over the individual response values in the control condition).

Counterfactual Thinking

The idea of an individual causal effect analysis is partially based on counterfactual thinking. Counterfactual thinking enters when we realize that we can only measure an individual's response to a single condition of a study. The thought process involves considering hypothetical information formed through counterfactual thinking. Counterfactual thinking involves hypothesizing the outcome of an event that did not happen, wondering what would have happened if things had been different.

Counterfactual thinking enters the causal effect analysis when considering a person's potential responses to multiple conditions of a study (Dawid, 2000). We use counterfactual thinking in everyday life when we hypothesize what would have happened if we had done things differently, or if circumstances had been different (e.g., *If I had just left my house a bit sooner, then I could've caught that subway train!*). In the context of RCM, an individual participates in a single condition of a study and we use counterfactual thinking to hypothesize about the response of this same individual in the other condition(s) of the study (Dawid, 2000). For the child in the violence prevention program above, if we imagine this child was actually in the treatment condition, we would engage in counterfactual thinking when considering how the child would have responded to the outcome measures if she had been in the control condition instead.

Counterfactual thinking also applies to the average causal effect. The average causal effect entails not only observed responses from participants in a study. It is

calculated, in theory, from both observed and hypothetical responses of individuals, with the hypothetical responses deriving from counterfactual thinking. The average causal effect can be thought of as the average of the individual causal effects. As the individual effects result from observed and hypothetical responses, so does the average because it consists of these individual effects. Social scientists apply the results of studies to the behaviors and attitudes of groups as well as individuals. Using counterfactual thinking, the value of the average causal effect can also be inferred as the individual causal effect, on the basis of certain assumptions.

Assumptions

Assumptions in Rubin's Causal Model refer to conceptual and methodological ideas that cannot be fully tested empirically and so must be assumed as true in order to make a valid causal inference given that each participant cannot be measured in all conditions. In writings about the model, a few assumptions are discussed. Not all assumptions receive the same focus in each discussion, and each discussion does not mention all assumptions. The assumptions of (i) independence (ii) the constancy effect and (iii) stable unit treatment value assumption appear in the main literature on RCM.

Independence

Independence (Holland, 1986a; Rosenbaum, 1984a) is an extension of the random assignment concept. A main tenet of random assignment is the way in which participants are allocated to conditions, i.e., that the responses participants provide are in no way related to biasing characteristics of the individual participants.

The assumption of independence focuses on the relationship between the condition one is exposed to and the value of the responses one would potentially supply

to all conditions in a study. Independence refers to the lack of association between the *set of responses* a participant would supply to conditions and the way in which participants are allocated to conditions. For example, person A is a participant in a study and is randomly assigned to the treatment condition of the study. If the response that person A would supply in the treatment condition and the response that person A would supply in the control condition are in no way related to the fact that person A was assigned to the treatment as opposed to the control condition, this would support the independence assumption.

The emphasis of independence lies in the potential responses a participant could supply. If this set of responses is not conditional upon exposure to conditions, independence is satisfied. Consider the idea that if participants in the treatment condition were instead exposed to the control condition, and participants in the control condition were instead exposed to the treatment condition, the causal effect estimate should not differ depending on which group is exposed to which condition. If the estimate does differ depending on which group is exposed to each condition, assignment to conditions is not independent of responses.

Independence may be satisfied through assignment of participants to conditions (e.g., random assignment, matching) or through the idea of strongly ignorable treatment assignment (Rosenbaum, 1984a)³, which relies on potential responses being independent of assignment to conditions based on covariates. Later in this paper, during the

³ If a researcher conducts a nonrandomized study, and variables related to the formation of the groups used in this study can be measured, these variables can be used as covariates in the study. With the use of these covariates, the validity of causal conclusions drawn from such a study is strengthened. This is the idea of 'strongly ignorable treatment assignment,' and has been formulated by Rosenbaum & Rubin (1983) and Rosenbaum (1984a).

explanation of Table 4, an example of a nonrandomized study is used to illustrate RCM notation; the importance of independence is further explained at that point.

Independence is satisfied by the idea of randomization, but is not identical to randomization. While randomization guarantees independence (to the limits of probability), other mechanisms can result in independence. Independence differs from randomization because it takes into account the observed and hypothetical responses to conditions. Although after being introduced to this way of thinking it seems to make intuitive sense in relation to random assignment, taking observed and counterfactual responses into consideration is not part of how we typically conceptualize random assignment.

The assumption of independence supports the use of the average causal effect in lieu of the individual causal effect, which is impossible to obtain. If sets of responses are independent of condition allocation, this is true for the group as well as the individual, and therefore the leap from inferring the average causal effect to inferring the individual causal effect comes naturally.

Constancy Effect

A second assumption that aids in inferring the individual level causal effect from the average causal effect is the ‘constancy effect’ (Holland, 1986a), also labeled as additivity (Holland, 1986a) and treatment-unit-additivity (Cox, 1992). The constancy effect assumes that the value of the causal effect is the same for every participant. That is, exposure to the treatment condition of the study, for every participant, is assumed to result in a response that differs by a constant amount from the response that each participant would have supplied in the control condition. Upon comparing the difference

in responses obtained after each participant's exposure to the treatment condition and each participant's exposure to the control condition, this difference in responses for each individual (i.e., the causal effect) would equal a constant value (i.e., the same value for each individual).

This assumption of constancy is not without some controversy or confusion in the RCM literature. Dawid (2000) argues that this assumption is not easily testable. Cox (1992) posits that if an interaction results from a study, then the constancy assumption is violated because the treatment effect is not the same for all participants in the study: the effect of the treatment depends on another variable.

Similarly, Holland (1986a) explains that the assumption "can be partially checked..." (p. 949). By dividing a sample into subpopulations and estimating the average causal effect for each subpopulation, one can check to see if the causal effect is constant over these subpopulations. The constancy effect assumption lacks support if the effect varies over these subgroups; if the causal effects for these groups do not show variation, the plausibility of the assumption remains. In Holland's (1986c) response to Cox (1986), he states that if no variables are measured in addition to the outcome, then there is no way to "falsify" the constancy effect assumption. In a way, the idea of the constancy effect speaks to the generalizability of a causal effect.

Stable Unit Treatment Value Assumption

The third assumption discussed in the RCM literature is the stable unit treatment value assumption, labeled as SUTVA. SUTVA contains two main aspects. The first aspect of SUTVA insures that the treatment and ensuing response to treatment is independent of the assignment mechanism implemented in a study. As Rubin writes,

“SUTVA is simply the a priori assumption that the value of Y for unit u when exposed to treatment t will be the same no matter what mechanism is used to assign treatment t to unit u and no matter what treatments the other units receive...” (1986, p. 962). That is, whether participants were randomly assigned to ‘treatment A,’ or were already exposed to ‘treatment A’ through the existence of natural groups, the outcome resulting from exposure to ‘treatment A’ should be the same under both conditions for any particular unit.

This is not to say that SUTVA is about implementing a treatment in a stable way, that the treatment is implemented in the same way for each participant. It is about assuming that the treatment is not unknowingly represented by multiple versions, slightly differing from one another, without notice to the researcher drawing causal conclusions about this seemingly ‘invariable’ treatment.

The second aspect of SUTVA assumes a lack of interaction among participants in a study. SUTVA indicates that the experience of one participant does not impact the experience of another participant in a study, thereby avoiding interference with the responses participants provide in a study. These two aspects of SUTVA become important when we shift our thinking from the causal effect for an individual participating in a study and providing potential responses to all conditions, to the average causal effect, which consists of responses from multiple participants. The inclusion of multiple participants raises these concerns about versions of treatment and interference among participants. When SUTVA is met, these issues no longer remain a concern.

Summary

Under the assumptions of independence, the constancy effect, and SUTVA, the average causal effect can be interpreted as the effect of the treatment variable for each participant in the study. Statistical analysis of data almost always involves comparing average responses under various conditions, not the responses of separate individuals. Ultimately, the level of analysis for the average causal effect yields an analysis of causal effects *from only observed data*, which can then, through the use of these assumptions and counterfactual thinking, be applied to individuals exposed to a single condition (as will be illustrated in an upcoming example).

Illustrating the Causal Effect Analysis Using Notation

RCM uses mathematical symbols and notation to express the tenets of the model. These symbols and notation will be defined and then used to illustrate an example of causal effect analysis. See Table 1 for a summary of notation for variables in the model. The main concepts represented in the notation, which have been discussed in detail above, include 1) manipulating exposure to levels of a variable, 2) attribute variables, 3) covariates, and 4) potential responses:

Exposure: S

The idea of exposing participants to conditions has been prominent throughout the above discussion of Rubin's Model. In the Model's notational system, the variable 'S' indicates the condition to which a participant is exposed. For example, if a researcher is conducting a study with only 2 conditions, a treatment and a control condition, then S would hold only two possible values: treatment (t) and control (c). For example, $S = t$ would indicate exposure to the treatment condition; $S = c$ would indicate exposure to the control condition. An 'S' variable is one for which participants can be potentially

exposed to multiple values. The values of an 'S' variable indicate the levels of a causal variable.

Subpopulation indicator: G

Another type of independent variable in the model is defined by the letter 'G.' Variables defining preexisting groups, "such as race or gender," (Holland & Rubin, 1983, p.6), fall under the category of 'G' variables in Rubin's Model. "Unlike S, it is not possible for the value of G for each unit to have been other than what it is" (Holland & Rubin, 1983, p. 6). 'G' variables therefore match the description for attribute variables. Attribute variables include variables for which a person can only be exposed to a single level; the value a person reports for an attribute variable cannot be other than what it is. The level of G a participant belongs to is indicated by $G = x$, where 'x' equals a value of the G variable. For example, if G represented sex of participants, $G = m$ would represent men participants, and $G = w$ would represent women participants.

In his later work, Holland (1986a) does not include the 'G' notation in his explanation of the model. Instead he presents 'A' to represent attribute variables, which are the variable types relevant for associational, not causal, studies. His description of 'A' variables seems quite broad, including any attribute and not only categorical variables, as 'G' seems to indicate.

Covariate: X

'X' represents a covariate in the model. A covariate is a type of independent variable, but a covariate cannot be affected by the treatment variable if it is to be a useful control. 'G' variables qualify as a type of covariate; however, 'G' differs from 'X' in that 'X' can be a causal, as well as a continuous, variable.

Response variable: Y

The variable ‘Y’ symbolizes a response (i.e., outcome). There are as many versions of ‘Y’ as there are conditions of an ‘S’ variable. The condition to which a response belongs is represented by a subscript attached to ‘Y.’ For the following discussion, the subscript ‘t’ refers to the treatment condition, and ‘c’ refers to the control condition. Response from a person in the treatment condition is symbolized as ‘Y_t’; a response from the control condition is symbolized as ‘Y_c.’ The causal effect for an individual is $Y_t - Y_c$, which is the difference in this individual’s response to the treatment and control conditions. Whether a response is an observed or counterfactual response, all potential responses are represented by ‘Y.’

Additional notation for the model

Expected values.

The average causal effect refers to the difference in the average response of participants in each condition. The model refers to the average response to each condition as the expected value of ‘Y’ (i.e., response) for each condition. “The expected value of a statistic is the mean of its sampling distribution” (Glass & Hopkins, 1996, p. 393). Therefore, the mean value of a variable in a condition is the response value that an “average person” would be expected to supply in that condition. (This mean value may not be a possible value: e.g., if half of the people in a study have a causal effect of 1, and the other half have a causal effect of 3, the expected value is 2, even though no individual in the sample has a value of 2.)

The letter ‘E’ symbolizes the expected value (i.e., mean) of a variable. RCM expresses the mean response for a particular condition in a study as the expected value of

the response for that condition, and symbolizes it using, for example, ' $E(Y_t | S = t)$,' which, in words, reads: The expected value (E) of the response (Y) to the treatment condition (t) for those individuals exposed (S) to the treatment condition (t).

In RCM, the comparison of responses from conditions is, ultimately, a comparison of expected values. The comparison of the expected value for the treatment condition with the expected value for the control condition results in the average causal effect estimate. Through counterfactual thinking, the model assumes that every person has a possible response value for each condition. [Note: Expected values refer to populations, as opposed to samples, of people.]

In Rubin's Model, when assuming independence, (which can be met through random assignment), even though every person only participates in the one condition s/he is assigned to, the expected values for the treatment and control condition equal the hypothetical scenario of *the expected value had every participant actually produced a response for each condition*. That is, the average causal effect resulting from comparing the expected value of the treatment group response (i.e., the mean response value if everyone had been exposed to the treatment) with the expected value of the control group response (i.e., the mean response value if everyone had been exposed to the control) includes an observed and a hypothetical value from each individual in the study. This will be demonstrated in the example below. [Note: Expected values are not observable features, just as counterfactual responses may not be observable.]

Conditional expectations.

The use of a conditional expectation in RCM arises from the need to distinguish between (i) an expected value calculated *solely from observed responses* supplied by

participants exposed to a condition and (ii) an expected value calculated from *both observed and hypothetical responses* (i.e., observed responses from participants exposed to the condition and hypothetical counterfactual responses from participants not exposed to the condition). To represent the expected value of the response from a particular condition, given that the average being represented is based *solely* on the participants exposed to that particular condition, a conditional expectation is used. A conditional expectation communicates the fact that a value is based on a specific condition – in the upcoming example, that condition will be group assignment.

An example of the type of conditional expectation used in RCM is the following: $E(Y_i|S=t)$. In words, this conditional expectation represents ‘the expected value of the response for the treatment condition (Y_i) conditional upon exposure to the treatment condition ($S = t$). [Note: the vertical line is interpreted as ‘conditional upon’ or ‘given that’ to indicate the boundary of inclusion in the statement.]

Table 2 shows the responses from four randomly assigned participants in the running example of the hypothetical violence prevention program with two conditions: a treatment condition and a control condition. Each of these four children participated in a single condition. Joseph and Carol were exposed to the treatment condition; Jennifer and Steven were exposed to the control condition.

In Table 2, a bold number represents an observed response for the condition the participant actually participated in, and the number not in bold represents the hypothetical counterfactual response, meaning the response the person would have provided had s/he instead participated in the other condition. For example, Joseph was exposed to the treatment condition, and provided a response of ‘3’; he was not exposed to the control

condition, but had he been, his response would have been '5.' This exemplifies the use of counterfactual thinking in the model.

Table 3 illustrates the use of RCM notation in this example. As discussed above, RCM uses causal effect equations to calculate the value of a causal effect in a study. Here are the expected value equations for the causal effect using the data in Table 3 (which is the same data as presented in Table 2):

$$E(Y_t) = (\mathbf{3} + \mathbf{4} + 3 + 4)/4 = 3.5 \quad (\text{Expected value for treatment condition response}) \quad (1)$$

$$E(Y_c) = (5 + 6 + \mathbf{5} + \mathbf{6})/4 = 5.5 \quad (\text{Expected value for control condition response}) \quad (2)$$

$$E(Y_t - Y_c) = E(Y_t) - E(Y_c) = 3.5 - 5.5 = -2 \quad (3)$$

To interpret equation 3 in words:

The average causal effect =

Expected value for treatment response – Expected value for control response)

These equations consist of the observed responses (in bold) and the hypothetical responses (not in bold inside the parentheses). The expected value of the causal effect (equation 3), is, therefore, based on the idea that every participant in the study has provided a response for each condition in the study, one observed response for the condition actually participated in, and a counterfactual response for the other condition. Because every participant in the study cannot be exposed to every condition, the researcher actually only observes:

$$E(Y_t | S=t) = (3+4)/2 = 3.5 \quad (4)$$

$$E(Y_c | S=c) = (5+6)/2 = 5.5 \quad (5)$$

Equation 4 represents the conditional expected value of the response in the treatment condition; this value is calculated only from the responses of those participants *who were actually exposed* to the treatment condition (which means excluding responses for those who were only hypothetically exposed to the treatment condition). Equation 5 represents the conditional expected value of the response in the control condition calculated only from the responses from those participants who were actually exposed to the control condition.

Understanding the *difference* between equations 1 and 2, and equations 4 and 5, is crucial to understanding the contributions of Rubin's Model. Equations 1 and 2 contain both observed and hypothetical values, while equations 4 and 5 contain only observed values. When will equations 1 and 4 and equations 2 and 5 be equal; that is, when will

$$E(Y_t) = E(Y_t | S=t)$$

and

$$E(Y_c) = E(Y_c | S=c)?$$

Or, said another way, when will equation 3, the average causal effect including both observed and hypothetical responses to each condition, be equal to the average causal effect resulting from only those exposed to the treatment or control condition:

$$E(Y_t - Y_c) = E(Y_t) - E(Y_c) = E(Y_t | S=t) - E(Y_c | S=c) \quad (6)$$

One situation in which the expected values calculated from observed and hypothetical responses (equations 1 and 2) will equal the expected value calculated only from observed responses (equations 4 and 5) is when the independence assumption has been met. One way to meet this assumption is through using random assignment. Random assignment ensures that participants assigned to be exposed to the control condition, for example, do not differ (on average) from participants assigned to be exposed to the treatment condition. Therefore, with random assignment, the expected value of the response from the treatment condition for all observed and hypothetical responses will equal the expected value of only the observed responses.

To demonstrate the relationship among random assignment, the expected value calculated from observed and hypothetical responses, and the expected value calculated from only the observed responses, consider that participants in the violence prevention study were not randomly assigned to conditions. The data from this hypothetical study appear in Table 4. In this study, Joseph and Jennifer were exposed to the treatment and Carol and Steven were exposed to the control condition.

Notice that for each person, the causal effect is $Y_t - Y_c = -2$.

According to these data:

$$E(Y_t) = (3+4+3+4)/4 = 3.5 \quad (7)$$

$$E(Y_c) = (5+6+5+6)/4 = 5.5 \quad (8)$$

$$E(Y_t | S=t) = (3+3)/2 = 3 \quad (9)$$

$$E(Y_c | S=c) = (6+6)/2 = 6 \quad (10)$$

In this nonrandomized study, equations 7 (which consists of observed and hypothetical responses) and 9 (which consists only of observed), and equations 8 and 10, are not equal. For the study reported in Table 4, preexisting groups were used and perhaps the participants exposed to the treatment condition were older. Perhaps older children respond to both the treatment and control conditions with lower values than younger children would. The data in Table 4 exemplify a situation in which the independence assumption is not met through the common method of random assignment. But conditional independence can be met through the use of covariates. The fact that the older participants were selected (or self-selected) into the treatment condition makes the rationale for exposure to conditions dependent on age. Following the idea of strongly ignorable treatment assignment (Rosenbaum, 1984a), if the variable 'age' is measured and statistically controlled for, responses can be made conditionally independent of exposure.

Unlike the randomized study in Table 2, had the participants in conditions in Table 4 been reversed, meaning that had Joseph and Jennifer been exposed to the control and Carol and Steven been exposed to the treatment, the value of equation 9 and of equation 10 would have been different than they are. If the participants were to be rearranged in this way, equation 9 would read, $E(Y_t | S=t) = (4+4)/2 = 4$; equation 10 would read $E(Y_c | S=c) = (5+5)/2 = 5$. In such a situation, we would incorrectly conclude from the observed responses that the causal effect was -1, not -2.

Regardless of how participants are (or are not) assigned to conditions, the unconditional and conditional expected values of responses are always related through the following equations (Holland & Rubin, 1983):

$$E(Y_t) = E(Y_t | S=t) P(S=t) + E(Y_t | S=c) P(S=c) \quad (11)$$

$$E(Y_c) = E(Y_c | S=c) P(S=c) + E(Y_c | S=t) P(S=t) \quad (12)$$

These equations weight the responses to conditions by the proportion (P) of people in each treatment condition represented by each conditional expectation. For example, in equation 11, the expected value of the treatment response $[E(Y_t)]$ equals the expected value of the response to the treatment condition for those participants exposed to the treatment condition $[E(Y_t | S=t)]$, weighted by the proportion of people exposed to the treatment condition $[P(S=t)]$, added to the expected value of the response to treatment for those participants exposed to the control condition $[E(Y_t | S=c)]$, weighted by the proportion of people exposed to the control condition $[P(S=c)]$. Although the only observable pieces of equations 11 and 12 are $E(Y_t | S=t)$, $E(Y_c | S=c)$, $P(S = t)$, $P(S = c)$, these equations demonstrate the underlying rationale for the relationships. These equations demonstrate that $E(Y_t) - E(Y_c)$ does not necessarily equal $E(Y_t | S=t) - E(Y_c | S=c)$.

Criticisms of Rubin's Causal Model

Analysis of Causal Effects

Rubin's Causal Model offers specific criteria for studying causal relationships.

The main tenets of Rubin's Causal Model are based on fundamental elements of research

design. But the model expands the way we think about some of these elements in such a way that helps us to identify and elucidate omissions and implicit assumptions in a study.

For example, the central counterfactual idea of the model, that when we analyze causal results from a study we are in theory considering each participants' responses to all conditions, is not an established way of conceptualizing data, although it does appear in general discussions of research methods. The idea that the causal effect calculated from observed responses when randomization is used equals the causal effect calculated from both observed and hypothetical responses (i.e., potential responses) is thought of as obvious by Leamer (1988). According to Leamer, this would mean that the idea of independence does not differ from the idea of randomization, and is therefore unnecessary. However, the assumption of independence can be met without random assignment, using the idea of strongly ignorable treatment assignment (Rosenbaum, 1984a).

Random assignment conveys the idea that participants randomly assigned to groups are considered interchangeable (but not necessarily equal). This differs slightly from thinking about people as potentially being exposed to each condition of a study; the idea of random assignment does not lead one to conceptualize a missing data problem as RCM does. Independence takes the randomization idea one step further, taking into account the full set of responses a participant would supply to all conditions in a study.

While some may say this is implicit in the idea of randomization, this idea does not appear in discussions of randomization. The conceptual bases of RCM may seem intuitive, but perhaps this hindsight bias (also common when one learns of a social psychological theory) is the mark of a well-conceived model. Keeping the set of

responses in mind, rather than assignment to a condition depending on chance, reminds us that each individual has a value for a causal effect, not that just a sample has a value (i.e., the causal effect estimate).

Dawid (2000) offers a critique on the role of counterfactual thinking in Rubin's model, expressing doubt as to how we would interpret the validity of conclusions inferred from results that cannot be empirically tested. This doubt is based on questioning causal inference when using 1) counterfactual thinking about what individual participants *would do* in alternative conditions and 2) untestable assumptions (such as the constancy effect). Dawid states that these ideas contradict the principle of 'falsifiability' in scientific research.

However, criticizing the model at this level is akin to questioning whether or not the principle of falsifiability itself is falsifiable. Methods are constructed as frameworks within which we test hypotheses. The conclusions supporting or not supporting hypotheses must be falsifiable. If we apply the falsifiability principle to the method itself, how would we ever be able to test this?

The counterfactual approach assumes that a person supplies a specific response when in the treatment condition, a different response when in the control condition and that the difference between these two responses is the individual causal effect. Hypothesizing that a person would react in a specific, predetermined way to a particular treatment has been criticized as fatalistic and deterministic (Dawid, 2000). In everyday social science research, though, researchers take the results of studies and apply these results to the populations appropriate for generalizability. The same critique could be targeted at using statistical analyses based on averages and generalizing such results to

populations of individuals. It is not assumed that all people act identically, or that their behaviors are predetermined. Researchers need to utilize methodological and analytical tools in order to draw conclusions about samples, and to ultimately understand individual behavior.

The link between counterfactual thinking and cause-effect relations is not particular to the RCM model. Shadish, Cook & Campbell (2002) acknowledge the role of counterfactual thinking in causal analysis back in the 18th century. Extant literature in the political science discipline (e.g., Fearon, 1991), and empirical research in the attitudes and social cognition areas of the social psychology discipline contributes to the general field of counterfactual thinking. In everyday life, people commonly use this type of thinking when trying to make decisions and consider the consequences of actions (e.g., Spellman & Mandel, 1999). Tetlock & Belkin (1996) emphasize how ingrained counterfactual thinking is in our understanding of causality by stating that if we want to avoid counterfactual thinking, we would be left relying solely on observed events, and would be unable to use the host of words, such as ‘influence’ and ‘precipitate,’ used to imply causal relationships in our everyday thinking. [See Roese (1997) for a review of the literature in this area.]

In response to Dawid’s criticisms, I believe that the RCM equations offer a realistic bridge between what a researcher can do in practice to reach causal conclusions, and what a researcher can consider hypothetically to infer causal conclusions. The idea of distinguishing between what can be done in theory and what can be done in actual research practice has been discussed by Blalock (1961a). He focused on a

theoretical/empirical language distinction: The idea of what could be done in reality versus what is only possible in hypothetical terms.

Applying Blalock's idea to the RCM model, the theoretical language could refer to 1) basing causal effects on counterfactual, hypothetical outcomes that could not be observed for an individual, and 2) the definition of a causal variable (i.e., what a person could possibly have been exposed to). The empirical aspect(s) of this model are 1) the observed outcomes, and 2) the assumptions that can be readily tested given the observed outcomes. Holland & Rubin (1983) and Wainer (1991) have applied the RCM model specifically for the purpose of identifying untestable assumptions in the research designs of studies.

For example, in their 1983 chapter, Holland & Rubin analyze Lord's Paradox. Lord's Paradox (Lord, 1967) is a situation in which two statisticians analyze the same data from a nonrandomized study, but each statistician's results lead to a different conclusion about the data. To summarize, the data these statisticians analyze come from a study of college students who are exposed to a cafeteria diet. All students are exposed to the same diet, but the statisticians attempt to 1) analyze the effect of this diet and 2) determine if the diet had a different effect on male and female students. Holland & Rubin (1983) use the RCM framework to explicate the assumptions involved in analyzing this data when there was no control group. By making assumptions about the unobserved responses to the nonexistent control condition, Holland & Rubin turn descriptive conclusions into possible causal conclusions.

Shadish (in press) articulates a few potential weaknesses of RCM. He classifies this model as one to be applied after a study is carried out, signifying that it focuses on

post-data collection strategies rather than strong research design. A danger in this is researchers relying on post-data processes to correct for problems rather than research design to prevent problems. But Rubin's Causal Model need not be thought of only as something to consider after data have been collected. The model can be considered during the design of a study as well. For example, by thinking of a study in terms of RCM during the design stage, a researcher may notice which assumptions will become relevant after data collection and may want to measure additional variables to consider ways to test these assumptions.

Another weakness Shadish mentions is not a criticism of the model itself, but of the way in which it has been discussed in the scientific literature. This second criticism touches upon the complex ways in which assumptions of the model have been discussed in the literature, and the debatable nature of some assumptions. The technical language often used to explain the assumptions (such as independence and the constancy effect) expands the possibility of their being misunderstood. Cohen (1992) communicated this same general idea, speaking to the advantage of using nontechnical language when publishing on methodology and statistical issues in social sciences.

The potential exposability idea is the foundation of the causal effect and causal variable definition in RCM. The idea of a causal effect being analyzed through comparing potential responses is logical and fitting to the context of testing many social scientific theories. However, in the logic of this model, common social science research practices are questioned. Of utmost importance is the classification of attribute variables as non-causal variables. Concepts such as sex and race can be crucial in our research, and necessary to examine in many theoretical *and* methodological models.

Causal Variable Definition

The majority of this discussion about Rubin's Causal Model has focused on the causal effect equation and not the causal variable definition. This is largely because in the literature by the main proponents of Rubin's Model, the causal variable definition is not discussed in as much detail as the causal effect equation. This could be because the main articles and chapters about this model are found in statistics and economics journals, where the issue of classifying attribute variables as non-causal may not present such controversial issues as in the social sciences.

In the main RCM literature, small sections exist about the causal variable definition. An author usually presents the idea of what qualifies as a causal variable, supplies an example, and then this qualification becomes accepted, unless, as the critique below demonstrates, a comment in response to the article challenges the idea. In this section, I draw mainly upon social science literature to emphasize the critical role of 'attribute' variables in social science research.

Effects of Causes Versus Causes of Effects

A focus on the causes of effects highlights the action of an independent variable 'acting on' a dependent variable, with the concern mainly on the independent variable (those included and those not included in a study). A focus on the effect of causes highlights the outcome after an independent variable has 'acted on' a dependent variable. Focusing on the effects tilts the analysis to one of what happened, or what would happen if, whereas the focus on causes of effects turns ones attention to what lead to the happening.

At the very beginning of Holland's article (1986a) about Rubin's Model, he states, "The emphasis here will be on measuring the effects of causes because this seems to be a place where statistics, which is concerned with measurement, has contributions to make" (p. 945). But the discussion of what constitutes a causal variable is crucial to the analysis of causal effects. An analysis of the effects of causes cannot proceed with an analysis of the causes of effects. The fact that Rubin's Model supplies terms and conditions to classify causal versus attribute variables demonstrates this point.

In his reply to Holland (1986a), Glymour (1986) focuses on the causes of an effect, rather than the effects of causes, although it should be emphasized that most discussions of causes of effects and of effects of causes include attention to causes *and* effects. A discussion of Rubin's Causal Model cannot proceed without attention to the causes of effects.

In agreement with Holland (1986a), Glymour interprets a cause as an event, but he disagrees with the restrictions Holland includes in this interpretation of a cause. Using genetics as his example, Glymour proposes that for those variables seemingly devoid of a counterfactual condition as the result of an event, we identify the event at the beginning of the chain of events that led to this variable value. If we can identify the variable (the event) that lead to the occurrence of a response, then a person could potentially take on a different value of that variable, meaning that variable could actually be a causal variable (Glymour, 1986).

Glymour (1986) uses the example of Down's syndrome to exemplify the contribution that the level of analysis issue brings to this definition. He posits that if we backtrack to the actual event that 'caused' the attribute to be the way it is, we could

change this event and its outcome. If an intervention, such as genetic engineering, was thought of as an event that could change the genetic composition of a person, then it could be imagined that a person could or could not have Down's syndrome. The essential being of this person would still be embodied in him/her with or without Down's syndrome.

However, is a disability such as Down's syndrome an event or an attribute? A person affected by Down's syndrome can have numerous physical and social problems. It is not realistic to imagine that by changing a 'single' genetic attribute (i.e., the gene(s) causing Down's syndrome), numerous other aspects of the individual would not also change, defeating the purpose of changing the 'single' event causing this syndrome in the first place. Therefore, Down's syndrome falls back in the 'attribute' categorization of RCM. The example of Down's syndrome, then, brings us back into the circle of again trying to distinguish attributes from causes.

According to RCM, questions of causal inference are not relevant when the independent variable used cannot be considered a causal variable, and in order for a variable to be considered causal, it must fit the condition necessary for an analysis of causal effects: A participant must be "potentially exposable" (Holland & Rubin, 1983) to all values of the causal variable being used in the study. "For causal inference, it is critical that each unit be potentially exposable to any one of the experimental conditions. As an example, the schooling a student receives is an experimental manipulation in our sense, whereas the student's race or gender is not" (Holland & Rubin, 1983, p. 6). If a participant cannot take on values of a variable other than what s/he currently occupies, then this variable cannot be considered causal for RCM.

Another way to think about this definition is to ask the question, “Would it be possible for participants to be randomly assigned to conditions of this variable?”

Regardless of whether or not the variable is for a randomized or non-randomized study, if the answer is no, then according to this model, that variable cannot be a causal variable.

In his 1986(a) article, Holland recalls that he and Rubin once formed the motto, “No causation without manipulation” (p. 959).

According to this causal variable definition, the variable “stereotype threat” (Steele, 1997) is a causal variable. When a person belongs to a group for which a stereotype exists about performance on a particular task, that person’s performance on the task is affected by whether or not she is primed with the stereotype before performing the task. There is a stereotype that men perform better with math than women. When women are primed with this stereotype, they perform less well than men on a math test; when women are not primed, they do not perform less well than men (Steele, 1997). The variable “stereotype threat” can be classified as causal because a participant in a study of stereotype threat could hypothetically be exposed to the ‘stereotype threat’ condition or the ‘no stereotype threat’ condition.

Following this criterion attributes such as personality variables (e.g., self-esteem, anxiousness, extroversion), demographic characteristics (e.g., sex, race, and age), and even physical characteristics (e.g., height, physical attractiveness) cannot be considered causal variables. Changing an attribute such as a dimension of a person’s personality or a person’s sex, age or attractiveness would also change other aspects of the person.

Although this idea may be more difficult to imagine for a variable such as height, consider the social research relating status to attractiveness, e.g., the ‘what’s beautiful is

good' hypothesis (e.g., Dion, Berscheid & Walster, 1972). If others' perceptions of a person change based on that person's physical attractiveness, then changing a person's physical attractiveness certainly changes other things about that person.

Such variables are not causal because it is not clear what aspect of the variable is causing an outcome in any given situation (Rubin, 1986). As will be discussed later in this dissertation, Rubin's concern here involves the Validity Typology's concept of construct validity. RCM's inattention to this issue creates weakness in the model. In another article, Rosenbaum & Rubin (1984) explain that there is no sensible interpretation for manipulating a variable that cannot be altered in reality; a characteristic or attribute is a nonsensical manipulation.

What Perspective Can Social Science Offer the Causal Variable Definition?

An implicit belief in RCM is that attributes are independent variables, "The term independent variable is often used to refer to a treatment indicator variable like S , but it is also applied more loosely to include an entirely different type of variable. In our model this second type of variable is the subpopulation indicator variable G " (Holland & Rubin, 1983, p. 6).

Variables such as sex, race, and personality are practically always used as independent variables, but they are complicated independent variables. In many ways these variables can be thought of as dependent independent variables, as concepts shaped by environmental, societal attitudes and institutional policies (Fine, 1996), and therefore as manipulable constructs. When a difference between African American and White participants is found for an outcome, how is it interpreted? To what is this difference attributed?

Deciding that attributes cannot be considered causal variables influences the nature of conclusions drawn by social scientists studying behavioral and attitudinal differences between men and women, people of differing races, people with differing personality characteristics, etc. According to RCM, when attributes are used as independent variables, any conclusions drawn must be descriptive, not causal (Holland & Rubin, 1983). How do researchers *differentially interpret and apply* such descriptive versus causal conclusions? How do researchers and policy makers *differentially evaluate the importance* of a descriptive versus causal conclusion?

An interesting area of complication resulting from this causal variable definition arises in situations when researchers study sex or racial differences among participants. A number of social science (particularly social psychological) studies do include sex or race as a manipulated variable.

For example, the conception of sex as a causal variable is found in a study of individuals who were being evaluated for an academic job strictly on the basis of a curriculum vita (Steinpreis, Anders & Ritzke, 1999). The goal of this study was to assess whether or not sex of a job applicant affected how his/her curriculum vita was judged by faculty members at various universities. The experimental manipulation in this study was ‘sex of person whose c.v. was being judged.’ There were two treatment categories, man applicant’s c.v. and woman applicant’s c.v. Actual faculty members participated in the study, and were randomly assigned to evaluate one of two treatments: the c.v. of a man applicant or the c.v. of a women applicant. Participants in the two treatment groups received the exact same c.v., the only difference being the name on the c.v. (i.e., it was

either Brian for the man's c.v. or Karen for the woman's c.v.); these names were chosen to explicitly reflect the sex of the applicant.

In this study, the participants were potentially exposable to the man and woman conditions of the study because these conditions were defined by a characteristic of the c.v. reviewed by study participants (and not a characteristic of the participants themselves). All other information on the two c.v.'s was identical, allowing the researchers to attribute any differences found to the sex of the c.v.'s owner. If a situation is imagined in which we used an actual man and woman to test this same idea of sex discrimination in hiring practices, we would not be able to conclude that sex was the cause because many things, other than sex, would differ between this actual man and woman (e.g., personality, physical appearance).

In discussing the elements of Rubin's Model, Pratt & Schlaifer (1984) propose a situation in which sex can be manipulated, explaining that "Sex and age are manipulated in reality by assigning a job to a person of one sex and age rather than another." In response to this, Rosenbaum and Rubin (1984) express, "Unalterable characteristics of individual units, however, such as age or sex, are not treatments: age and sex do not have causal effects because it is not meaningful to discuss the consequences of altering these unalterable characteristics."

Considering the arguments for attributes not being considered causal variables, (e.g., Cox, 1992; Holland, 1986a; Holland & Rubin, 1983; Rosenbaum & Rubin, 1984), and considering the abundance of the type of study that Steinpreis et al. (1999) conducted, the interpretation (i.e., descriptive vs. causal) and applicability of conclusions from such studies is an important issue for social science researchers.

Social science defines common 'attribute' variables.

The question arises as to the current purpose and classification of attribute variables in social science. Is it 'established' knowledge that race and sex *can* or *cannot* be considered causal variables? What other variables fall under the attribute heading?

Personality.

Although no literature on Rubin's Model mentions 'personality' as a causal or attribute variable, psychologists often study personality as a factor. Researchers present differing views on how much our personalities are predetermined through genetics (e.g., Down's syndrome, intelligence, mental illness, sociability) versus shaped by environment (e.g., social class, early deprivation). A strong program of research backed by the American Psychological Association is currently underway to study links between specific genes and specific personality traits (Azar, 2002).

Personality has been defined as "...that pattern of characteristic thoughts, feelings, and behaviors that distinguishes one person from another and that persists over time and situations" (Phares, 1991, p. 4). This definition touches upon another personality research debate: Does personality consist of stable enduring traits or changing states? Within psychology there are debates as to the flexibility of one's personality – is personality an enduring trait formed in early childhood? Or is personality a changing entity into adulthood (Costa & McCrae, 1988)? Even the environmental approach to personality weighs heavily on the idea that personality changes occur most frequently in childhood, although it does not deny the fact that personality can change throughout the lifespan.

Of concern to the classification of personality as a causal versus attribute variable is its manipulability and the potential for an individual to be exposed to differing levels of personality. Personality research provides evidence that one's personality intricately relates to many aspects of a person's attitudes and behaviors, particularly in relation to health (e.g., Helgeson, 1994; Kobasa, 1979; Miller, Smith, Turner, Guijarro, & Hallet, 1996). For classification purposes in Rubin's model, given that other aspects about a person would change if a person's personality changed, personality would fall into the 'attribute' categorization, along with sex and race.

Demographics.

What would researchers of the evolutionary psychology discipline, who use sex as a causal variable, say about this model? Researchers in this area of psychology propose, for example, that over time men and women have developed different strategies to ensure the propagation of their genetic material (e.g., Buss & Schmitt, 1993). As a result, men and women, for biological reasons, look for different characteristics in a mate (e.g., Kenrick, Sundie, Nicastle & Stone, 2001). Sex, for researchers in this area of psychology, *is* causal – men and women act differently precisely because they are men or women.

In some areas of the social sciences, particularly those for which the study of sex/gender, race, and ethnicity are focal (e.g., the psychology of women) the complexities of these attribute variables are well known and recognized. But Zalk & Gordon-Kelter (1992) use the term “ghettoization of feminist studies” to convey how in research and teaching, issues of race, gender, and class often emerge as special topics rather than established, mainstream areas of study.

This leads to the question of how ‘mainstream’ researchers imbue these variables with meaning, and for what purpose they use these variables in research. The American Psychological Association’s fourth edition of its publication manual, in its instructions for how to describe research participants, reads, “Report major demographic characteristics such as sex and age. When a particular demographic characteristic is an experimental variable or is important for interpretation of the results, describe the group specifically...” (1994, p. 13).

The fifth edition of the manual expands, “Report major demographic characteristics such as sex, age, and race/ethnicity, and where possible and appropriate, characteristics such as socioeconomic status, disability status, and sexual orientation. When a particular demographic characteristic is an experimental variable or is important for the interpretation of results, describe the group specifically...” (2001, p. 18). The classification of demographic variables as “experimental variables” (no definition is supplied for the term ‘experimental variables’) implies that APA may classify demographic variables as causal. The APA manual’s reporting guidelines indicate 1) these variables can be considered ‘experimental variables’, and 2) the expectation that these attributes may influence the research results.

In these next paragraphs, only a basic discussion will be presented of the complexities involved in using variables such as ‘sex’ and ‘race.’ The multiple meanings and relationships among these variables go beyond the scope of this paper; the purpose of selecting these variables and discussing what issues and concepts they are used to represent must be considered by the researcher using these variables in a study. They will

be discussed here for the purpose of more clearly defining their roles in the context of the RCM model.

Race has been identified as a biological concept as well as a socially constructed group of categories (LaVeist, 1994). The variable ‘ethnicity’ is sometimes used to encompass race, as well as nationality (Phinney, 1996; Sue, 1999). Much disagreement exists among psychologists as to appropriate definitions of race and ethnicity (Phinney, 1996; Yee et al., 1993). Because of the political implications of race difference findings, and the lack of a clear conceptual definition of race, many psychologists tend to avoid the use of this variable in their research (Yee, Fairchild, Weizmann, Wyatt, 1993).

Sociologists also face debates on this issue. The American Sociological Association released a statement about the importance of studying race (ASA, August, 2002). This statement came in response to biologists’ and anthropologists’ opinion that the government should stop measuring race to prevent further social division in society. Sociologists, in return, emphasized the multitude of topics for which racial differences continue to be found. Race, they replied, remains a crucial concept in social categorization.

The variable ‘sex’ refers to the biological characterizations of men and women, based on physiological differences. The variable ‘gender’ is based on the idea that a person’s attitudes and behaviors are shaped by (i.e., caused by) social role expectations, societal influence, and stereotypes of men and women. Social psychologists (as well as researchers in other disciplines) regard sex and gender as different, but interrelated, variables (e.g., Deaux, 1993; Eagly, 1995), and there has been much debate about the

meanings of and distinction between these variables (e.g., Deaux, 1993; Laner, 2000; Unger & Crawford, 1993).

Proxy variables.

Researchers often use a variable to represent a construct or variable that was not actually measured in a study. Variables used to represent correlated but unmeasured variables are labeled as ‘proxy variables’ (e.g., Kaufman, 1997; LaVeist, 1994; Matthews, 1989; Walsh, 1999). In the literature, demographic label measures (Walsh, 1999) such as race, sex, and ethnicity are sometimes acknowledged as proxy variables, and are usually used as proxies. Often researchers do not consider the multiple meanings, markers and representations of an attribute variable as a proxy variable and use it at face value (LaVeist, 1994; Walsh, 1999).

Explicit discussions of the use of attribute proxy variables do exist in the health psychology and epidemiology literature. Although not explicitly labeled as proxy variables, discussions of the multiple constructs represented by measures such as ethnicity (e.g., Sue, 1999; Phinney, 1996), sex (e.g., Eagly, 1995; Laner, 2000) and race (e.g., Eisenman, 1995; Fine, 1996) appear in the general psychology and sociology (e.g., Laner, 2000) literature.

Race has been discussed as representing culture, discrimination, and biological differences (e.g. LaVeist, 1994). Fine (1996) interprets ‘whiteness’ as a “proxy for quality, merit, and advantage” (p. 38). Race has been used to represent power differences (McIntosh, 1993), class (Eisenman, 1995), “ethnicity, skin color, and nationality,” (LaVeist, 1994). Race also represents cultural differences in social perception, values, beliefs (Jones, 1997), norms and attitudes (Phinney, 1996), and daily experiences

(McIntosh, 1993). Pinderhughes & Shim (2001) reframe the problem of measuring race as a problem of assessing race relations and social processes.

Labels of ethnicity have been described as representing experiences related to one's status as a member of a minority. For example, discrimination and prejudice may be aspects related to one's status as a minority (Phinney, 1996). Sue (1999), while never explicitly using the term 'proxy variable,' expresses a very similar idea. He writes, "If ethnic differences are found, researchers should explain what aspects of ethnicity are responsible for the differences" (p. 1076).

Researchers commonly use 'sex,' which is a measure of biological difference, to represent socially constructed gender roles. Using a biological distinction to stand in for socially constructed roles classifies 'sex' as a possible proxy variable. There may be instances in which a researcher uses the variable 'sex' to test for differences in an outcome, when in fact gendered behavior/attitudes are actually being measured.

In the case of 'sex,' it is not possible to imagine a person being potentially exposable to other values of the variable (except in the case of surgery). In the case of gender, it is possible to imagine a person could be different than what they are. Social constructionist explanations support the belief that people can possess different qualities and abilities than they do. Revisiting Holland and Rubin's motto of 'no causation without manipulation' (1986a), we can imagine that gender roles could be assigned and manipulated. For example, theory by social constructionists (e.g., Eagly, 1995), supports socialization (which could, in theory, be manipulated) as a way through which gender becomes incorporated into one's identity. Starting in early infancy, people treat girl and boy babies differently (Seavey, Katz & Zalk, 1975), and young children are socialized to

play with gender-specific toys (e.g., Raaj, 1999). Different socialization patterns lead girls and boys to develop differing interests (e.g., Lindsey, Mize, & Pettit, 1997), abilities, and interpersonal communication styles (e.g., Fivush, Brotman, Buckner & Goodman, 2000; Tannen, 1990).

Because researchers often simply collect data on each participant's classification as a man or a woman, and use this as an independent variable, the meaning and interpretation of any relationships found between 'sex' and other variables remains ambiguous (Denmark, Rabinowitz, & Sechzer, 2000). This ambiguity also applies to differences found among racial categories. When a researcher reports a sex difference in findings, "so many uncontrollable variables are correlated with biological sex that we can never know which variable played a causal role in a gender difference." (Denmark et al., 2000, p. 14). This same confusion could be applied to differences found among racial categories. As Sherif articulates, "...the 'variable' called sex is like a railroad boxcar: everyone knows what it is called and what it is used for, but no one knows what is inside" (1994, p.124).

Proxy variables encompass multiple meanings and levels of complexity, including socially constructed categories often misused as essentialist, biological categories. For example, in the survey research literature, many researchers have reported that demographic indicators such as race, sex, age, and level of education predict a person's likelihood to not respond, or respond with a 'don't know,' to a survey item (e.g., Ferber, 1966; Francis & Busch, 1975). However, the more stable characteristic identified in this literature is education (Narayan & Krosnick, 1996; Witt, Pantula, Folsom & Cox, 1992). Race, sex, and age may actually be proxies for groups/categories of people to whom

education and exposure to the survey issues are not readily available, or for people who are not interested in the particular issues involved in the survey (e.g., there is a literature on ‘self-attributed ignorance,’ e.g. Turner & Michael, 1996).

As this proxy variable discussion highlights, often when researchers use variables such as sex and race to assess differences in an outcome, the actual idea conveyed entails variables *related* to the attribute variables (Walsh, 1999). Identifying these demographic labels themselves as the meaning of the variable minimizes the underlying complexities of the search, and leaves many unanswered questions about which causal mechanisms are actually impacting the variables under study. And as this discussion suggests, some variables are not easily classified as either causal or attribute variables. For example, socio-economic status is often used as a demographic indicator, or attribute. Is socio-economic status (SES), often referred to as ‘class,’ an event that happens to a person, or is it an attribute of a person?

It is possible to change a person’s income, or a person’s level of education, but in addition to income and education, socio-economic status relates to other things that build over time. For example, class relates to access to education, which relates to the type of job one is able to pursue, which will relate to income, quality of life, physical and mental health (Dohrenwend, Levav, Shrout, Schwartz, Naveh, Link, Skodol, & Steuve, 1992), and quality of life for one’s children, creating a multitude of variables related to, and even confounded with, SES.

SES could be considered an attribute in Holland’s (1986a) sense because many aspects of life may change considerably if a person’s SES were to change. But SES can change; it is not a biological or internal characteristic. Beyond the face validity of SES as

a variable, a measure of SES can be used to represent variables that define it. While SES per se may not be a causal variable, variables related to SES may be. For example, access to educational opportunities, health services, housing, and institutional policies may be causally affecting the outcomes of a study in which a relationship is found between SES and a dependent variable. SES is often measured through reports of annual income. While manipulating one's annual incomes is possible, annual income is only one aspect of SES.

Consequences

The use of attribute variables presents complex challenges. But simply deeming these variables as non-causal may limit their use and importance in social science research because, according to RCM, only descriptive conclusions may be inferred about attributes. The application of descriptive conclusions may be more difficult to interpret and implement, subsequently leading to their under-utilization by researchers and policymakers alike. This is not to say the information offered by descriptive conclusions is not useful information. For example, descriptive information may be used to predict outcomes.

In a way, RCM's causal variable definition guides researchers away from committing the fundamental attribution error. The fundamental attribution error (Ross, 1977) refers to the fact that people, particularly in the Western world (Semin & Zwiern, 1997), tend to attribute the causes of events internally, or to disposition, rather than context. For example, if a person fails a test, the fundamental attribution error would be committed when people attribute the failure to the intelligence level of the person, rather than, perhaps, the environment in which the test was taken, or the difficulty level of the test. The role of contextual causes is often overlooked. By not allowing attributes to be

interpreted as causal, RCM guides researchers away from attributing causes to variables often thought of as ‘internal’ to the person (e.g., sex, race), to variables outside of the individual that can be changed.

On the other hand, this definition of a causal variable complicates the use of what RCM classifies as attribute variables in research. These complications arise in different ways. For example, Pepitone (2000) asserts that a focus on manipulation of the independent variable prevents research on topics such as culture. “Culture is simply not like an experience or a stimulus; it is inextricably within individuals, brought by them into the experimental situation” (Pepitone, 2000, p.243). Returning to Glymour’s discussion of what constitutes a causal variable, would RCM classify culture as an attribute or causal variable? It can be imagined that at the most basic level, a person’s culture could be different than what it is, if that person had been raised in a different environment with different customs and beliefs. In principle, we could imagine a study in which children were randomly assigned to be raised in different cultures. Later in life, though, changing a person’s culture could potentially change numerous other things about the person.

Rubin’s Causal Model ventured to supply a definition for causal variables and classify attributes such as sex and race as non-causal. But this may simplify the complexities of these variables that are so crucial to social science research and remove them from discussion altogether, as often happens in the RCM literature.

CHAPTER 4: ELABORATION MODEL

The Elaboration Model originated in the discipline of sociology. Sociologists typically pursue research questions involving variables that cannot be manipulated for logistical or ethical reasons. They study issues as they naturally occur in society. Mainly for these reasons, sociologists do not conduct experiments and are unable to eliminate many possible alternative explanation variables through research design. Among other methods, sociologists, as well as psychologists and political scientists, conduct surveys. It was through the process of analyzing survey data that sociologists developed the Elaboration Model.

When designing a survey, a researcher has the opportunity to collect a wide range of information beyond the main variables of interest. This information can be used in the statistical analysis stage: After testing for a relationship between the main variables of interest, additional variables can be included in tests to elaborate on the original relationships of interest. When a relationship between two variables is observed, the Elaboration Model can be used as a guide to explore third variable explanations. The theory underlying the EM pertains to understanding a bivariate relationship at a deeper level of theoretical and statistical analysis, through multivariate analysis. Through multivariate analysis, the researcher can understand the original relationship at a more specific level.

Although one might wonder if the Elaboration Model must be applied to the study of causal relationships specifically, or if the model can be applied to elaborate on any type of relationship, in one of the few recently published pieces on the Elaboration Model, Babbie (2001) describes applications of the model in causal language. Rosenberg

(1968) limits the use of the Elaboration Model to “asymmetrical relationships” in which one variable X is “responsible for” (Rosenberg, 1968, p. 9) another variable Y.

Rosenberg describes six types of asymmetrical relationships, explaining that not all of them are causal. [Note: An analysis of the language used in Rosenberg’s explanations of these six types of relationships may lead one to interpret most of them in the causal realm.] Rosenberg (1968) and Babbie (2001) provide causal examples to explain the concepts of the model. It seems, though, that the model would apply to descriptive and predictive relationships as well as causal relationships given the model’s basic purpose of gaining a more thorough understanding of a bivariate relationship through the introduction of additional variables.

Babbie (2001), who has written much on the elaboration model, explains it in the following way: “The basic steps in elaboration are as follows: (1) a relationship is observed to exist between two variables; (2) a third variable is held constant in the sense that the cases under study are subdivided according to the attributes of that third variable; (3) the original two-variable relationship is recomputed within each of the subgroups; and (4) the comparison of the original relationship with the relationships found within each subgroup provides a fuller understanding of the original relationship itself” (p.426).

Test Variables and Partial Relationships

The third variable Babbie discusses is known as a *test variable* in the EM. The introduction of a test variable increases “one’s confidence that there is a meaningful and inherent link between the two variables” (Rosenberg, 1968, p. 37). Relevance of the test variable to the bivariate results, and the statistical relation of the test variable to those factors in the bivariate relationship guide the choice of test variables; one should make

sure that the test variable relates statistically to the independent and dependent variables before entering it into the analysis (Rosenberg, 1968). When a test variable enters the analysis, the data are further categorized into the levels of the test variable. The researcher examines data from the whole sample, but the data are subdivided according to the levels of the test variable. For each level of the test variable, the data are examined separately. Relationships examined according to the classification levels of a test variable are known as *partial relationships* (or conditional relationships). It is at the level of partial relationships that the researcher begins to form a deeper understanding of the original bivariate relationship under study.

For example, imagine that a sociologist wishes to investigate a relationship between the role of religion in one's life and age, based on findings that as one grows older, religion becomes a more significant factor. According to prior literature she read suggesting the role of biological sex in one's ideas about religion, she might be interested to know if the relationship between these two variables differs between men and women. Biological sex would then be entered into the analysis as a test variable, and data for the bivariate relationship between religion and age would be analyzed separately for men and for women. The original bivariate relationship between religion and age would then be analyzed at the level of the partial relationships, created by the levels of the test variable: men and women. The Elaboration Model is designed to then answer the following question: Does the relationship between the role of religions in one's life and age in men differ from the relationship between the role of religion and age in women?

The addition of a test variable, which is biological sex in the above example, creates a route of further understanding for the researcher to explore. Unlike the idea of

statistical control, the EM does not follow a specific statistical model. It can be used in combination with a variety of statistics: contingency tables, regression analysis, analysis of variance, etc. The EM presents a logic researchers can apply when examining a possible cause-effect relationship in nonexperimental data.

Originally, the EM was developed in terms of contingency tables. The purpose of using contingency tables to test the EM differs slightly from typical uses of these tables (Lazarsfeld, Pasanella & Rosenberg, 1972). Contingency tables, also known as cross-tabulations, are typically used for: 1) descriptive purposes - meaning investigating differences, which are often demographic, among subgroups in a sample (e.g., Are there more older people than younger people reporting poor health), and 2) investigating whether or not associations between variables exist (e.g., Do people who report drug use also report positive attitudes towards drugs?). Using the EM, a researcher starts with an established bivariate relationship and uses contingency tables to probe into why this relationship may exist, or specifically where in the sample it does exist. The contingency table set-up provides information about how two variables relate. Frequency data in each cell of the contingency table provide this information.

For example, returning to the question of religion and age above, imagine the sociologist studying this question administered a survey to 100 people, 50 young adults and 50 older adults. Table 5 presents the frequencies resulting from this hypothetical survey.

Table 5 demonstrates that 20% of young people felt that religion was important to their lives, and 80% of older people placed importance on religion. A chi-square statistic is usually used to test the relationship between the cells in a contingency table. A chi-

square answers the question: Are responses to the column variable independent of the row variable, or are responses to the column variable dependent upon which level of the row variable one is in? The significance level of the chi-square indicates whether or not a relationship is believed to exist in the table. For this table, the chi-square would be significant, indicating that the importance of religion in one's life depends upon age.

Table 6 illustrates the resulting partial relationships when a test variable is introduced, showing how the test variable would partial the sample into subgroups. The test variable, located above the dependent variable in Table 6, separates the data into two sub-tables, each table a test of the original bivariate relationship. The difference between the two tables is in the level of the test variable under study. The test variable of interest in this example is biological sex.

Examining the age-religion relationship separately for men and women shows no difference in the original bivariate relationship by sex. Therefore, this test variable is unnecessary to include in further analyses. But it has provided something for our understanding of the original bivariate relationship – that the relationship between religion and age is the same for men and women.

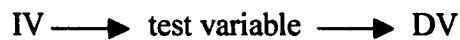
Temporal Order

The temporal placement of a test variable in relation to the row (independent) and column (dependent) variables classifies the resulting partial relationships into one of four types. Two types of temporal orders are relevant for the EM: Intervening relationships and antecedent relationships. Not all relationships must be intervening or antecedent, though; the model does account for relationships in which the test variable neither intervenes between the IV and DV nor precedes the IV and DV. But classifying the time

order sequence of a test variable supplies information about the causal order of variables in a relationship (Rosenberg, 1968).

Intervening relationship

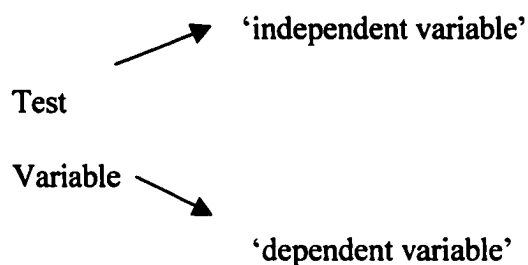
A test variable can intervene between the independent and dependent variables:



The addition of an intervening variable contributes to the understanding of why changes in the independent variable may cause changes in the dependent variable. In this chain of variables, the independent variable causes changes in the test variable, and the test variable causes changes in the dependent variable. More than one intervening variable may work in any given chain; given the impossibility of testing all relevant test variables, a researcher cannot assume that because one test variable intervenes in a relationship, that no other test variables may also fit into the pattern. An intervening variable is classified as such through theoretical reasoning and logic, not statistical analysis.

Antecedent relationship:

A test variable may also precede the original independent and dependent variables in temporal order:



In the case of the antecedent test variable, there are really two 'dependent' variables because both occur after the test variable. The test variable causes both the independent and dependent variables in an antecedent relationship. In such a case, the addition of a test variable indicates that the independent variable is not directly related to the dependent variable, as the original bivariate relationship would suggest. Introducing the antecedent variable allows the researcher to demonstrate that the relationship between the IV and DV was due to a third variable.

Classifying Partial Relationships

The temporal placement of a test variable in relation to the IV and DV comprising the original bivariate relationship contributes to the classification scheme of partial relationships. Table 7 (Babbie, 1992) classifies four types of relationships according to where the test variable falls in the theoretical chain of time, *replication*, *specification*, *explanation* and *interpretation*.

When the test variable occurs before the independent and dependent variables (i.e., when the test variable is antecedent) and its inclusion in the analysis explains away the relationship between the independent and dependent variables, the resulting partial relationship is classified as *explanation*. In explanation, the original relationship is

weakened because the inclusion of the test variable creates a stronger relationship with the dependent variable than the original IV did. An important omission in the model is a guideline for determining what constitutes a stronger or weaker relationship when the partial relationship is compared to the original bivariate relationship (Babbie, 2001).

When the test variable intervenes between the independent and dependent variables (i.e., an intervening test variable), and the resulting partial relationship is not as strong as the original bivariate relationship, *interpretation* results. Both explanation and interpretation signify that the independent variable is not related to the dependent variable conditional on the test variable.

Regardless of the temporal placement of the test variable, when the addition of a test variable does not significantly alter the relationship between the independent and dependent variables, the resulting partial relationships are labeled *replication*. Finally, when the relationship between the original two variables differs depending on the level of the test variable the resulting partial relationships are labeled *specification*. In accordance with Table 7, each type of resulting partial relationship will now be described through examples.

Deciding the temporal placement of a test variable involves considering issues such as common order of events (e.g., parent education occurs before child education) or the need for some variables to occur before others (e.g., deciding age, sex, and race takes place at birth) (Carrington & Dembski, 2001). Order of variables may be considered from the point of view of “properties” (Rosenberg, 1968). “Properties” are those variables such as race, sex, age, nationality, religion and eye color and are considered

temporally prior to other types of variables because “they are either strongly or absolutely resistant to influence” (Rosenberg, 1968, p. 17).

To demonstrate the four types of partial relationships, an example used in Lazarsfeld et al., (1972) and Babbie (1992) will be described. The example comes from a study conducted by Stouffer in the 1940s (e.g., Stouffer, 1972); Stouffer’s attempt to find alternative explanations for his unexpected results later led to the development of the Elaboration Model by Lazarsfeld and Kendall. Stouffer studied men in WWII for the purpose of better understanding what affected soldier morale. One of his hypotheses was that men with less education would not resent being drafted into the war as much as men with more education. In this hypothesis, level of education is the independent variable, which is assumed to be causally prior to the dependent variable of attitude about being drafted. Stouffer’s data did not support this hypothesis.

Stouffer used social psychological theory to understand why, and he concluded that possibly attitudes toward being drafted were affected by whether or not one’s friends were also drafted. Whether or not one’s friends were also drafted became the test variable. Using this example, the four types of partial relationships presented in Table 7 will be described.

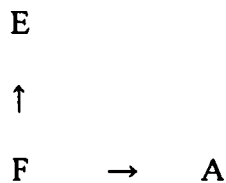
Replication

Once a test variable enters the analysis, the original relationship, as defined by the original independent and dependent variables, becomes partialled according to the levels of the test variable. If the partial relationships replicate the original IV-DV relationship, the test variable proves unnecessary in the analysis, but still offers a new piece of information (i.e., that this specific variable does not affect the bivariate relationship). For example, in Stouffer's study, if the relationship between education and attitude remained the same for those whose friends were and were not drafted (i.e., the original relationship was replicated in the partial relationship), we would rule out friend's draft status as a possible reason for the results.

Explanation

If, once the test variable becomes part of the analysis, the original IV-DV relationship disappears then this test variable has helped identify a spurious relationship and becomes the explanatory factor in this original relationship. Imagine that we tested the original education/attitude relationship within the subgroup of soldiers whose friends were not drafted, and we observed that the relationship between education and attitude was not statistically significant. Then we also found a nonsignificant relationship between the original two variables (education and attitude) for those whose friends were drafted.

We would conclude that the original relationship between education (E) and attitude (A) was spurious, meaning it could be explained away by the inclusion of the third variable: whether or not one's friends were drafted (F). In the following diagram, the arrows represent the direction of relationship between variables.



This diagram indicates that the differences in education were not causing differences in attitude - whether or not one's friends were drafted, because whether or not a person was drafted depended on their level of education.

Interpretation

If the test variable does not explain away the original relationship but, rather, elucidates why the original bivariate relationship exists, this test variable has aided in interpreting the original relationship. For example, the original relationship between education and attitude about induction can be better understood when we know whether or not the inductees' friends were deferred. If an attitude is related to education only when deferment status is the linking variable between education and attitude, then the resulting partial relationship can be labeled as interpretation. Babbie writes, "...an interpretation does not deny the validity of the original, causal relationship but simply clarifies the process through which that relationship functions" (1992, p. 421).

Interpretation seems to be similar to the concept of mediation. In mediation, the relationship between two variables exists through the presence of a third. In this example, the relationship between education and attitude exists through whether or not one's friends were deferred. In interpretation, the test variable must be intervening: $E \rightarrow F \rightarrow A$. Notice in this diagram that a causal relationship from E to F is nonsensical: One's education cannot cause one's friends to be deferred. Friends' education relates to their being drafted (again, it seems a stretch to say that friends' education causes their

deferment). This example indicates that the relationships relevant for the EM must not always be causal relationships. Here, education can predict friends' deferment.

Here, the test variable 'friends' leads to the classification of "interpretation" for the resulting partial relationships because it assists the researcher in interpreting the relationship between the original two variables.

Specification

The fourth scenario is specification: when the test variable specifies under what condition, or for what subgroup the relationship holds. Specification occurs when the original bivariate relationship for one subgroup of the test variable differs from the bivariate relationship in the other subgroup of the test variable. For example, if, once friends' deferment enters as the test variable for the bivariate relationship between attitude about induction and education, the relationship between the original two variables only holds for those people whose friends were deferred, this would classify the partial relationships as specification.

The idea of specification seems to match the concept of moderation (e.g., Baron & Kenny, 1986) indicating an interaction between the independent variable and the test variable (a second independent variable). In an interaction, the relationship between the independent and dependent variables depends on the level of a third variable. The third variable, or test variable in this case, specifies under what condition the original relationship holds.

These four types of partial relationships communicate some of the ways that a test variable can clarify a bivariate relationship. The partial relationships allow the researcher to better understand possible causal relationships and draw more accurate conclusions

about the conditions of a causal, as well as non-causal, relationship. Rosenberg (1968) also suggested two types of relationships beyond these four, suppression and distortion.

Suppressor and distorter relationships

When a researcher finds a bivariate relationship of zero, the relationship may still be worth investigating further. Within the bivariate relationship may be a third variable complicating the detection of the relationship between the first two variables. In 1968,

Suppressor.

A suppressor variable is one that may hide or weaken the existence of a relationship between two variables. In such a case, a researcher may find that the relationship she expected to find did not appear in the results. The more common reaction to such a scenario is to believe the results (perhaps questioning the method or sample) rather than search for a suppressor variable.

Imagine a hypothetical study in which length of time foreign-born citizens had been living in the United States and attitude towards a war between the United States and Iraq were the two variables of original interest. The researcher finds no relationship between these two variables. The variable 'age' is then considered. It appears that younger people are more supportive of the war effort than older people, but younger people have not been living in the U.S. for as long a time as the older people. When the variable 'age' is added to the analysis as a test variable, a relationship is found such that among older people, those who had been living in the United States the longest were most supportive of the war efforts.

In this example, age is a suppressor variable, obscuring the relationship between length of time in U.S. and attitude toward the war. Without measuring age and including

it as a test variable in the analysis, we would not have discovered the relationship between time and attitude that existed contingent upon age. A suppressor relationship can be thought of as the inverse of the partial relationship called 'explanation.' Whereas with explanation a relationship is originally found between the IV and DV, but disappears upon the introduction of the test variable, with suppression, no relationship exists between the original IV and DV but is uncovered upon the introduction of the test variable (Rosenberg, 1968).

Distorter.

In the four basic types of partial relationships discussed above, the consideration of partial relationships being weaker than the original relationship is included in the discussion of specification. To cover the case in which the partial relationship may actually be the reverse of the original relationship, Rosenberg (1968) proposes the idea of a distorter variable.

Using the scenario in the suppressor variable example above, imagine a positive relationship exists between attitudes toward war with Iraq and length of time in U.S. such that the longer a person has lived in the United States, the more supportive she feels towards war with Iraq. Suppose age is added as a test variable. The relationship of attitude with time in the U.S. within age categories actually shows that the longer a person has lived in the U.S. the less likely she is to support a war with Iraq. The addition of age as a test variable produces the reverse relationship between the original two variables.

CHAPTER 5: VALIDITY TYPOLOGY

Researchers in psychology, education, medicine, and other disciplines seeking knowledge of causes and effects are well versed in the typology of validities, first developed by Campbell in the late 1950s. The earliest works on the typology included only two types of validity, internal and external validity. Cook & Campbell (1979) expanded the typology to include four validities, including construct and statistical conclusion validity. Shadish, Cook & Campbell (2002) present an even further developed explanation of the Validity Typology, providing a more advanced understanding of each validity's intricacies.

Because the VT model is well known in social science, for the purposes of this paper, it will be summarized, and then later explained in relation to the EM and RCM. The VT model is based on the idea of ruling out plausible rival hypotheses for the relationship between independent and dependent variables, and judging the degree of generalizability of the conclusions reached. Campbell & Stanley (1966), Cook & Campbell (1979), and Shadish, Cook & Campbell (2002) apply the idea of validity to assessing the truthfulness of causal inferences. They explicitly define validity as referring to the level of truth in a statement.

The Validity Typology consists of four types of validity: statistical conclusion validity, internal validity, construct validity and external validity; each type of validity carries with it a list of possible threats. These threats represent possible alternative explanations of results. If certain threats are plausible for a study, yet cannot be ruled out, they weaken the validity of conclusions drawn. By asking to what degree each of the

threats to validity can be ruled out, the researcher gains a better sense of the degree of accuracy of a causal conclusion.

Types of Causal Conclusions

There are two types of causal conclusions described by Shadish et al. (2002): causal description and causal explanation. To differentiate these types of conclusions, Shadish et al. distinguish between ‘molar’ and ‘molecular’ concepts; these two terms refer to the level of specificity inherent in a concept under study. The word ‘molar’ indicates a general concept that consists of many pieces; ‘molecular’ refers to the specific elements of which a molar concept consists.

Causal *description* refers to the general relationship between two variables: “molar outcomes and molar treatments” (Shadish et al., 2002, p.10). Causal *explanation* takes place at the “molecular” level, specifying what aspects of the treatment and outcome create the relationship between the variables.

Shadish et al. provide an example to illustrate this distinction: If a researcher were to conclude that psychotherapy (a molar treatment) leads to less depression (a molar outcome), this would be a causal description because we only know that these two general concepts, psychotherapy and depression, are related. To provide a causal explanation, we would want to know *what* about psychotherapy (i.e., which molecular pieces) leads to the decrease, and what specific aspect(s) of depression are affected. The causal explanation helps us understand the causal description.

While causal description is a strength of experimentation, causal explanation is not (Shadish et al., 2002). Experiments often lead to general conclusions about a cause and effect, without details of the mechanism working in the cause-effect relationship.

Causal explanation, which specifies aspects of the mechanism, aids in the process of generalizing results. Most experiments do not produce the level of detailed description required for causal explanation.

Causal Variables

Shadish, Cook & Campbell (2002) distinguish between causal variables that can be manipulated and those that cannot be manipulated. They *seem* to communicate the same idea as RCM, that attributes [“people’s ages, their raw genetic material, or their biological sex” (p.7)] are not manipulable and therefore cannot be causal variables in experiments. However, Shadish et al. go on to say, “To be clear we are not arguing that *all* causes must be manipulable – only *experimental* causes must be so. Many variables that we correctly think of as causes are not directly manipulable” (p. 8). Their reference to experimental causes can be translated as a manipulable cause – one that can be manipulated during an experiment.

In contrast, nonmanipulable causes are those that cannot be directly manipulated. Shadish et al., unlike Holland and Rubin, support the study of nonmanipulable variables as causes. “Nonetheless, nonmanipulable causes should be studied using whatever means are available and seem useful” (Shadish et al., 2002, p.8), the idea being that these studies will lead us to discover those variables that we can in fact manipulate.

According to this stance, experiments are not the only way to study causes, and limiting causation to manipulation greatly limits the types of questions researchers can answer. Rubin would support this idea. Through studying nonmanipulable causes, though, Shadish et al. assert that we can sometimes find manipulable variables that can be studied as representative of the nonmanipulable cause, and that can advance research on

the topic of the nonmanipulable cause. [Note that this idea raises the same issue that the proxy variable idea addresses.]

These authors also raise the issue of encountering an effect for which we wish to find a cause. Using the example of AIDS as an effect, they explain that experiments are not always possible to uncover causes, which Holland and Rubin clearly agree with, but where Shadish et al. differ is in wishing to find the cause of this outcome, whether or not that cause is one we can manipulate.

The difference in the stance of the Validity Typology's approach to classifying a causal variable and Rubin's Model's approach is striking. The two models greatly *differ* in this respect, leading a researcher to two possible inferences about any conclusion involving a nonmanipulable variable. If the researcher is using Rubin's Model as a guide, she will likely not deem the inference a causal one, but if she is using the Validity Typology as her guide, she may very well interpret the inference as a valid causal inference (depending on the absence of other threats to validity, of course).

Four Types of Validity

Being able to infer valid causal conclusions according to the VT requires considering specific threats to four types of validity. Each of the four types of validity is accompanied by a list of *possible* threats, but not all the threats are *plausible* in every study. A researcher needs to judge the plausibility of each threat on a study-by-study basis. "However, the plausibility of a threat is always contextually dependent on the joint characteristics of the design, on extra-study knowledge about the threats, and on the pattern of observed study results" (Shadish et al., 2002, p. 139).

Statistical Conclusion Validity

Statistical conclusion validity concerns the validity of inferences about the relationships among variables. Threats to statistical conclusion validity concern possible reasons why a researcher may not detect covariation among variables or may detect an incorrect degree of covariation. Examples of threats include low statistical power, violating assumptions of statistical procedures, and performing multiple tests without correcting for the significance level. If these, or any of the other threats in this category, present a plausible problem, confidence in the statistical conclusion validity of a study and the study's conclusions, is weakened.

Internal Validity

Internal validity refers to the degree to which we can be sure that the observed covariation between variables in a study can be attributed to the independent variable(s) of interest. Threats to internal validity involve alternative explanations for why a relationship may be found between an independent and dependent variable. Internal validity is the type of validity most relevant to questions of causal inference.

To increase the internal validity of a study, it is necessary to rule out additional variables that may be influencing the study results: "Accounting for third variable alternative interpretations of A - B relationships is the essence of internal validity..." (Cook & Campbell, 1979, p. 50). Here, one can see a similarity between the Elaboration Model and the Validity Typology given that both focus on considering third variable explanations.

Examples of threats to internal validity include whether or not the participants were selected in a way that makes the study groups unequal in some respect before

exposure to the independent variable, the uniformity with which the measures are implemented and collected, and the pattern, if any, noticeable for participant attrition.

Construct Validity

Construct validity pertains to how well the concepts of interest, meaning the “persons, setting, treatment and outcome” (2002, p. 65) were translated into variables. How a researcher labels and carries out the theoretical ideas of a study affect how the study results will be interpreted and applied. Correctly labeling concepts is crucial to construct validity. Examples of threats to construct validity include poorly defining a variable, using only one method to operationalize a variable, the influence of the experimenter’s expectations, and resentment of a control group due to the realization that other study participants are receiving things that they are not.

External Validity

External validity, like construct validity, applies to persons, settings, treatment and outcomes. The issue of external validity is generalizability – to what extent can study results be generalized to other people, conditions, treatments and measured outcomes beyond the study? Threats to external validity include, for example, not being able to generalize results about a causal relationship across different classes of units (e.g., does the relationship hold for both males and females?), and across different settings. Confusion sometimes arises when distinguishing between construct and external validity. Construct validity pertains to generalizing *to* causes and effects; external validity pertains to generalizing *across* causes and effects (Shadish et al., 2002).

Campbell, Cook, Shadish and Stanley discuss in detail many different types of research designs, describing possible validity problems that may arise within each design.

They clearly state that what is being judged is the validity of the inferences drawn from a study, not the study design itself. But some designs do lend themselves to strengthening specific aspects of validity. For example, the optimal way to rule out many threats to internal validity is to conduct a randomized experiment. However, although this may increase the internal validity of a study, it may result in decreased external validity because the controlled conditions of a randomized experiment limit the generalizability of results.

Notation

To acknowledge aspects of a study's design that relate to strengths or weaknesses for specific threats to validity, Campbell & Stanley (1966) developed a system of notation to illustrate the structure of designs. The notation does not help a researcher understand validity, but does indicate basic elements of research design indicating certain threats that may be more plausible than others.

The VT uses symbolic language to represent study designs. A study design can be represented through the use of two letters: X and O. 'X' indicates the 'experimental treatment,' that a group was exposed to a treatment. 'O' indicates an observation, i.e., a data collection point. When these symbols are presented horizontally, they are to be interpreted as occurring in a temporal sequence. Each horizontal sequence represents a single group; multiple groups are represented by a vertical succession of sequences. When a study includes two groups, these can either be depicted as separated by a dashed line or not (Shadish, 2002). The dashed line indicates that a comparison group was used, but not created by randomization. When this line is absent, randomization is indicated. In

addition, the letters 'NR' preceding the X, O sequence for each group indicates random assignment did not take place. For example:

1) A one-group pretest-posttest design is represented by:

O X O

2) A posttest only control group design is represented by:

X O

O

A dotted line signifies that a comparison group has not been "equated by randomization"

3) For example:

NR X O

O

Groups not equated by randomization might be age groups, ethnicities, or classrooms.

By depicting the structure of the design in these ways, threats to validity may become more apparent. For example, by looking at diagram three, we see that the control group was not created through a randomization process. Therefore, a plausible threat to the internal validity of this study might be how the participants were selected because it could be characteristics of the participants, rather than 'X,' that may lead to differences in 'O.'

The VT has been applied to at least 16 different research designs, experimental and quasi-experimental, demonstrating the widespread utility and applicability of judging validity of inferences in this way (Campbell & Stanley, 1966).

CHAPTER 6: REGRESSION

Chapter 2 of this dissertation concentrated on the use of research design and statistical analysis to rule out third variable explanations. The importance of accounting for third variable explanations guides various components and perspectives of the three methodological models I have just discussed.

Rubin's Causal Model focuses its concern about third variable explanation on the assignment mechanism used by the researcher to allot participants to conditions. Rubin (2002) comments that in many research articles, no details are supplied for how participants ended up in conditions. Through independence and SUTVA he emphasizes the importance of specifying the assignment mechanism in order to acknowledge possible assumptions made on the part of the researcher, and for the appropriate interpretation of causal inferences.

The Validity Typology presents a more well-rounded perspective on the third variable matter. The VT approaches the issue of control from a research design standpoint by thoroughly considering the multitude of circumstances and events that could introduce alternative explanations into the research scenario.

The Elaboration Model approaches controlling for third variables from yet another perspective. The EM does not so much attempt to control for third variables; instead it encourages the inclusion of additional variables to assist the researcher in developing a more thorough answer to the research question. The types of partial relationships resulting from including a test variable provide the researcher with a deeper appreciation for the intricacies of the causal relationship under study.

A main method of statistical analysis when aiming to account for third variable explanations is the regression framework. Regression will be considered here because it is the method of choice for so many social scientists using quantitative analysis. It is the type of statistical analysis social scientists typically use to explicate relationships among variables, and it is the basis of other analyses such as analysis of variance, analysis of covariance, and structural equation models that researchers use to analyze causal relationships. In the leading academic journals of the personality and social psychology field, the use of multiple regression has increased since the 1970's, from being used in 5% or fewer published articles in 1976 to being used in over 40% of articles in 1996 (Sherman, Buddie, Dragan, End & Finney, 1999).

Regression may be viewed as a fourth approach to causal analysis, but recall that RCM, the VT and the EM are not statistical models, they are methodological models. A number of statistical analyses, including regression and analyses based on regression, can be applied to evaluate causal relationships and effects for any one of those three methodological frameworks. In this section the three methods models are not to be compared to the regression framework. Instead regression will be explained here because its logic often guides the choice of and interpretation of results for causal questions.

Statistical models do not speak to the validity of a causal conclusion. Statistical models and significance levels do not distinguish between valid causal inferences and questionable causal inferences. The validity of a causal inference is not solely determined by the statistical analysis chosen for the data; it is also greatly dependent upon the research design and data collection choices (Licht, 2000).

Although the premise of regression analysis is based on prediction (Spirtes, Glymour, & Scheines, 1993), it is often implicitly interpreted in a causal way. A distinction has been identified between two purposes of regression analysis: applied prediction and theoretical explanation (Licht, 2000). Cohen & Cohen (1983) state that while correlation does not equal causation, “Causation manifests itself in correlation, and its analysis can only proceed through the systematic analysis of correlation and regression” (p. 15). They suggest that when conducting a hierarchical regression analysis, one should enter the independent variables in order of “causal priority.” This often leads to first entering variables such as “age, sex, ethnicity, education, and socioeconomic status” (Cohen & Cohen, 1983, p. 121).

The long held notion that demographic predictors should be entered first in regression might not be the most appropriate approach to such analysis. Instead, theoretical reasoning should guide the entry of variables into an analysis; the researcher should determine the entry order of variables according to where they fall in the theoretical chain of events.

Recall that analysis of variance is a specific case of the broader regression framework. The analysis of differences between an experimental and control group could be studied either through an ANOVA or a regression. Depending on the research design and control variables included, one would likely infer a causal conclusion after analyzing the data with ANOVA. This same data could be analyzed in a regression equation, and should render the same causal conclusion as the ANOVA. The general framework of regression fits well with the analysis of causal effects because testing causal models

through regression allows the researcher to ‘isolate’ the variance in the outcome due to the treatment when the appropriate control variables have been included in the model.

Choosing Control Variables

The statistics field does not provide much guidance for deciding which variables should be controlled for (Boruch & Rindskopf, 1977). There are not many restrictions on which variables can and cannot be used as control variables (Cohen & Cohen, 1983). The decision of when to control for variables should depend on the theorized causal order of the variables (Blalock, 1961b) and should be made before carrying out a study (e.g., Cochran & Rubin, 1973) so that relevant variables can either be taken care of through assignment of participants or by measuring possible confounding variables to be used in statistical analyses. Many researchers decide on the criteria for including a covariate on a “case-by-case basis” (Pearl, 1997). The problem of choosing control variables has been described as one of “most frustrating issues in causal analysis” (Pearl, 1997, p.4).

Two issues arise when considering the use of control variables. The first deals with how confounding variables might affect possible causal inference regarding the effect of the independent variable on the dependent variable. Being able to draw conclusions based on considerations of alternative variables strengthens the internal validity (Campbell & Stanley, 1966; Cook & Campbell, 1983). Controlling for variables is a case of testing alternative hypotheses, making sure that a relationship other than the one of interest is not responsible for the outcome. Having tested alternative hypotheses, a researcher increases internal validity and gains more confidence in the causal conclusions being drawn (Campbell & Stanley, 1966; Cook & Campbell, 1983; Rindskopf, 2000).

Given that the purpose of controlling for a variable is to rule out an alternative cause, and given that Rubin's Model does not classify nonmanipulable variables as causal variables, does RCM restrict what can be used as a control variable? Rubin (1974), Cook & Campbell (1979), and Cohen & Cohen (1983) use mostly causal language when discussing the use of control variables. Rubin (1974) discusses the importance of the researcher having a model for the relationships being tested. In order for a researcher to be able to adjust for a control in a meaningful way, the researcher must have some idea of the "functional form of the causal effect of these other variables on Y" (Rubin, 1974, p. 696). Blalock (1961b) suggests controlling for those variables "assumed to be causally prior" to the independent variable of study.

Interpreting the relationship between the control variable and the outcome as causal provides a reason for including it in a causal model. Controlling for a variable related to the outcome increases the precision of the causal effect estimate. Rubin writes, "In both randomized and nonrandomized studies, the investigator should think hard about variables besides the treatment that may causally affect Y and plan in advance how to control for the important ones..." (1974, p. 700). How does Rubin's model settle the discrepancy between the fact that control variables are causal, 'G' variables are a type of control, or covariate, but 'G' variables cannot be causal?

Removing the effect of a control variable highly related to the outcome helps a researcher conclude that differences in the outcome are not a matter of differences in this variable. "It is of the utmost importance that the investigator carefully think through the causal implications of the meaning of 'Y from which covariates have been partialled,' and 'equating groups on covariates' to assure himself that they make substantive

theoretical sense” (Cohen & Cohen, 1983, p. 425). A researcher must carefully analyze the logic behind controlling for attributes related to the actual groups of study (Cohen & Cohen, 1983).

A second issue arising in the relationship between control variables and causal inference involves the relationship between the independent variable of interest and the control variables. Controlling for variables related to the independent variable decreases bias in the estimate of the treatment effect. A variable affected by the treatment should not be used as a control variable unless the effects of the treatment on this variable are slight in relation to the actual treatment effect; removing the variance accounted for by such a variable would mean removing part of the treatment effect in the outcome (e.g., Cochran, 1983; Rosenbaum, 1984b). If random assignment has been used in the study, then no relationship should exist between the independent variable and the control variable, and bias is not a concern. However, in a non-randomized study, bias is a primary concern because, for example, variables related to how groups were formed could be biasing the statistical results.

Statistical Models

Multiple regression and analysis of covariance are two common statistical methods used when hypotheses about control variables are tested. Analysis of covariance fits within the more general framework of regression (Cohen & Cohen, 1983). When running a regression analysis or an analysis of variance model with more than one independent variable, the parameter estimate for each variable in the model is statistically controlled for the other variables in the model, whether we wish to use any of the variables as control variables or not. The estimate of each parameter is calculated to

reflect the variance that a variable contributes to the outcome over and above the other predictor variables in the model.

When a researcher wishes to control for a specific variable to rule it out as an explanation for an outcome, the analysis of covariance is often chosen as the statistical method if the independent variables are categorical. Multiple regression is often chosen when the independent variables are continuous, although it can handle categorical variables as well if they are properly coded.

Earlier in this paper, the idea of statistical control was discussed in conceptual detail. But in statistical terms, what does it mean to say that a variable has been controlled for? Entering a control variable into a statistical model (e.g., a regression equation) indicates that the variance it explains in the outcome has been accounted for. The following regression equations will be used to illustrate this point:

$$\hat{Y} = b_0 + b_1 X_1 \quad (13)$$

$$\hat{Y} = c_0 + c_1 X_1 + c_2 X_2 \quad (14)$$

Equations 13 and 14 are two possible ways of analyzing the data from a study. In both equations, Y is the response variable and X_1 is the continuous independent variable of interest; b_1 , c_1 , c_2 are the regression coefficients. In equation 13, no additional independent variables have been placed in the equation; no variables are being controlled for. If there are other variables that should have been controlled for, then b_1 , the estimate of the effect of X_1 (which is the independent variable whose effect we are interesting in knowing) will be biased. It will not be an accurate representation of the effect of X_1 on Y .

In equation 14, X_2 is a confounding variable the researcher wishes to control for. If the omission of X_2 from equation 13 had been biasing the result, then controlling for X_2 would mean that the variance in Y explained by X_2 has been accounted for, and separated from the variance in Y explained by X_1 . Therefore, the coefficient c_1 is the estimate of X_1 on Y controlling for X_2 . Said in another way, c_1 is 'the effect of X_1 on Y for people who are equal on X_2 ' (i.e., conditional on X_2).

c_1 is assumed to be a more accurate estimate than b_1 because the possible biasing effects of X_2 have been removed from it. But how would we know if c_1 is in fact a more accurate estimate than b_1 ? Under what circumstances will b_1 (i.e., the effect of X_1 on Y) equal c_1 (i.e., the effect of X_1 on Y controlling for X_2)? When no relationship (i.e., correlation) exists between X_1 and X_2 , omission of X_2 would not bias the results – there would be no 'omitted variable bias' (Rindskopf, 1986). If b_1 equals c_1 , and c_1 is accurate, then b_1 is unbiased, and the lack of a control variable in equation 13 was not biasing the results. In such a case, if X_2 and Y were significantly correlated, X_2 could still be controlled for to increase precision (decrease error variance).

In the more specific context of comparing groups in an analysis of variance framework (Cochran, 1971), imagine a study with one independent variable, X_1 . This variable has two conditions, control and treatment. In the following equations the control and treatment conditions will be represented by the numbers 0 and 1, respectively. $\bar{Y}_1$ represents the average response from the control condition and $\bar{Y}_2$ represents the average response from the treatment condition. Equations 15 through 17 exemplify the analysis without a covariate.

$$\bar{Y}_1 = b_0 + b_1(0) = b_0 \quad (\text{the control group}) \quad (15)$$

$$\bar{Y}_2 = b_0 + b_1(1) = b_0 + b_1 \quad (\text{the treatment group}) \quad (16)$$

$$\bar{Y}_2 - \bar{Y}_1 = (b_0 + b_1) - b_0 = b_1 \quad (\text{treatment} - \text{control}) \quad (17)$$

Therefore, b_1 is the estimate of the treatment effect when no covariate is included in the equation. It is the unadjusted treatment effect. If a covariate, X_2 , was added to the equation, $\bar{X}_{21}$ would be the mean of the covariate in the control condition and $\bar{X}_{22}$ would be the mean of the covariate in the treatment condition. The analysis of covariance would then proceed as follows, in equations 18 through 21:

$$\bar{Y}_1 = c_0 + c_2 \bar{X}_{21} \quad (\text{control group}) \quad (18)$$

$$\bar{Y}_2 = c_0 + c_1 + c_2 \bar{X}_{22} \quad (\text{treatment group}) \quad (19)$$

$$\bar{Y}_2 - \bar{Y}_1 = c_1 + c_2 (\bar{X}_{22} - \bar{X}_{21}) \quad (20)$$

$$\bar{Y}_2 - \bar{Y}_1 - c_2 (\bar{X}_{22} - \bar{X}_{21}) = c_1 \quad (21)$$

Now, the adjusted treatment effect is c_1 . The interpretation of c_1 differs from that of b_1 in that the effect of X_2 has been removed from c_1 . By controlling for X_2 , the mean difference on the control variable between the treatment and control conditions was linearly removed from the mean difference between the two groups on Y , the response variable. This subtraction of the adjustment from the response results in the adjusted

treatment effect, c_1 . The treatment effect is now adjusted for the covariate – the variance contributed by X_2 has been removed.

Structural Equation Modeling

An extension of the regression statistical framework often used to study causal relations is structural equation modeling (SEM); these models are also known as causal models. SEM builds upon the basic ideas of multiple regression analysis. Structural equation models consist of two pieces: a structural model of the relationships among latent, unmeasurable constructs, and a measurement model of the relationships among measured variables (Joreskog, 1977). Both pieces are built through and connected via a series of regression equations. The overall model supplies information (parameter estimates) explicating how observed (measurable) variables are related to one another as well as to underlying (unmeasurable) constructs.

Unlike regular regression, regardless of the research design, structural equation modeling techniques are interpreted as causal models, given that the structure of the model and the underlying theory are considered valid (Cohen & Cohen, 1983; Sobel, 2000). More generally, SEM is considered to provide a framework through which causal theories can be tested. If, through the series of equations, a SEM fits the data well, this would be consistent with a particular causal model. The use of SEM in social psychological and personality studies has greatly increased in recent years, from being used in 0% of published articles in the major journals in 1976 to 14% in 1996 (Sherman et al., 1999).

Important differences exist between the conceptual approaches of RCM and structural equation models. Structural equation modeling does not include criteria for

classifying variables as causal or not causal. SEM does not exclude attributes from being causal variables; sex, race, etc. are used as causal variables within this framework.

Therefore, some feel that distinctions have not been made in the SEM literature between meaningful and meaningless causal statements (Holland, 1986a). Tenets of RCM seem to clash with the concept of a control variable, though. If a control variable is a causal variable, how, by definition, can Rubin's Model classify 'G' variables (attributes) as a type of covariate?

In addition, the SEM framework does not place the same emphasis on manipulation of treatment variables that the RCM does. Whereas the analysis of causal effects in the RCM model depends on the outcomes resulting from various conditions of a manipulated variable, the analysis of causal effects in SEM can involve sets of relations of numerous variables across simultaneous equations, involving both measured and unmeasured constructs (Spirtes et. al, 1993).

A weakness in the use of SEM relates to the basis of the VT. Researcher using SEM rarely explore plausible rival hypotheses (Rindskopf, 2000). Conclusions from structural equation models are seldom considered in terms of alternative directions of relationships between variables, or with regard to variables not included in the model (Rindskopf, 2000).

Relationship with Other Models in This Paper

The regression framework can be thought of in concert with any of the three models (the VT, EM, RCM) already discussed because those three models do not in themselves contain a specified type of statistical analyses. Evaluating the way in which the data are analyzed affects the validity of causal conclusions because if data are

incorrectly analyzed, then conclusions will not be accurate (Shadish et al., 2002). Later in this paper, when the VT, EM, and RCM are applied to published research studies, the regression framework will also be considered in terms of how the researchers in those studies approached their analyses.

CHAPTER 7: COMPARING AND CONTRASTING RCM, EM, and VT

Three methodological models have been described above: Rubin's Causal Model, which offers a framework for aiding researchers in highlighting the key pieces in an analysis of causal effects; the Validity Typology, which offers researchers a way to judge the plausibility of alternative hypotheses and rule out plausible threats to the accuracy of their conclusions; and the Elaboration Model, which encourages researchers who have found a bivariate relationship to statistically test for the relationships created by adding a test variable(s) to the analyses.

The VT has been used to assess validity in psychological, educational and other areas of research. The Elaboration Model has been used mainly in sociological research. What perspective does Rubin's Causal Model offer existing social science methods for an analysis of causal inferences? To begin answering this question, these three models will now be compared and contrasted to identify their unique, as well as overlapping properties.

Defining a Causal Variable

On a critical point, these models differ with respect to how a causal variable is defined. RCM's definition has been explained and questioned extensively in this paper. Rubin's Causal Model includes a specific causal variable definition using the idea of potential exposability. Its criterion that a causal variable must be manipulable differs from the VT and EM. The EM literature makes no mention of a definition for a causal variable. In the context of the VT approach, Shadish et al. (2002) do explicitly state that causes can be manipulable or nonmanipulable. This distinction in how the models approach the discussion of causal variables can lead to a difference in opinion about what

constitutes a valid causal inference. Shadish et al., for example, unlike RCM, do not explicitly prevent causal conclusions from being inferred through the use of attribute variables.

Types of Studies and Relations

The VT distinguishes between two types of causal relationships: description and explanation. Description refers to a general relationship between two concepts, and is the goal of most experiments. Explanation refers to specifying the mechanisms at work in a relationship; this is not a strength of experiments.

RCM distinguishes between three types of studies: controlled causal studies, uncontrolled causal studies and descriptive studies. The first two types of studies include ‘S’ variables, and therefore causal conclusions; the third type of study involves only attribute variables, and therefore only descriptive conclusions.

The EM classifies four basic types of partial relationships, and two additional types of relationships (suppressor and distorter), all based on the placement of and resulting relationships from a test variable. Unlike RCM and VT, these partial relationships do not dictate the type or depth of conclusions inferred. Instead, the EM classifies relationships for the purpose of helping the researcher understand how the test variable furthers the theory under study.

Differing Foci

An analysis of the validity of causal inferences is, in the case of RCM, guided by being able to satisfy definitions, equations and assumptions, and in the case of the VT, by ruling out specific threats. The assumptions pertinent to an analysis using the RCM include independence (or ‘strongly ignorable treatment assignment’ in the case of

nonrandomized studies), the constancy effect and SUTVA. In the VT, the potential threats to consider are many. The EM does not attempt to estimate a causal effect value as RCM does, but there is an attempt to explain, empirically, what might be happening in the relationship between two variables. Those who utilize the EM may not even be interested in true experiments that result in a causal estimate value.

Rubin's Causal Model does not take into account many of the issues that the VT addresses. It does not consider issues such as construct validity, statistical power, or generalizability in the same sense that the VT does. But both of these models focus on the validity of causal inferences. From the perspective of the RCM, at the most basic level the validity of causal inferences entails abiding by the definition of a causal variable, and determining whether or not the requisite pieces for calculating a causal effect are present or can be estimated. From the perspective of the Validity Typology, the validity of causal inferences refers to whether or not plausible alternative hypotheses have been considered and ruled out.

Assumptions versus Threats

An understanding of RCM involves understanding the assumptions of the model (e.g., independence, constancy, SUTVA). The stable unit treatment value assumption (SUTVA) shares certain qualities with threats to construct validity in the VT. As described in detail in Chapter 3, the second aspect of SUTVA assumes a lack of interaction among participants in a study and therefore, a lack of interference among participant responses. This aspect of SUTVA resembles the threats of compensatory rivalry, resentful demoralization and treatment diffusion. In this regard, RCM does broach the subject of construct validity.

RCM uses the term ‘assumptions’; the VT uses the term ‘threats’. Perhaps the difference in approaches simply grew out of the difference in disciplines of origin. The VT originated in social science, where people are the participants and interactions among participants commonly occurs. This may relate to the model’s approach of considering and deciding whether or not to rule out threats. RCM originated in statistics, where participants of studies are several steps removed from the numbers they provide, and assumptions are easier to conceptualize and implement.

Acknowledging the Role of Theory

An important issue is Rubin’s Causal Model’s lack of taking theory into account. The EM paradigm is guided by the goal of theoretical elaboration. For example, in Stouffer’s study about soldier morale described above, the test variable of whether or not one’s friends were drafted was chosen through using social psychological theory to elaborate on the meaning of the original research. The Validity Typology considers theory when, for example, it concentrates on which elements of a research study have and have not been taken into account and controlled for when planning and carrying out the study, and the degree to which theoretical constructs have been operationalized accurately.

The Elaboration Model may be more helpful in terms of wanting to find the *meaning* behind results and evaluating the ‘truthfulness’ of this meaning. The RCM does not venture into theoretical meaning or including theory. It is concerned with the numbers and mathematical relationships, the structure of the study, not the meaning. RCM focuses on what Shadish et al. (2002) term causal description, the basic relationship between treatment and outcome. For example, Meyerowitz, Chaiken & Clark (1988) note that

there is an established relationship between the experience of a chronic illness, such as breast cancer, and sex-role stereotypes. But their interest is in what, specifically, about breast cancer and what, specifically, about sex roles is active in this relationship. Using this example, RCM would be able to speak to the validity of an inference about the basic structural relationship between breast cancer and stereotypes, but what it would be able to offer beyond that to help us understand this relationship is questionable.

RCM mainly concerns issues of internal validity (West et al., 2000). As Shadish et al. write, “Internal validity concerns an atheoretical local molar causal inference” (2002, p. 463). RCM largely ignores what Shadish et al. term causal explanation – the reasoning behind what about the treatment and what about the outcome contributed to the causal relationship. As discussed earlier in this paper, Rubin (1986) touches upon the idea of construct validity when he explains one reason why he does not consider sex to be a causal variable. But he does not expand upon this at all, leaving sex and other attribute variables out of further discussion in RCM.

Involving Mathematics

Rubin’s Model differs most obviously from these other two in that it is more of a mathematical model; its implementation is based on a set of equations and assumptions that must be fitted in order to assess a causal effect. The Validity Typology does not explicitly involve mathematics. Statistical concepts are used to determine the degree of statistical conclusion validity in a study, but numbers and equations are not necessary to implement this model. With regard to the Elaboration Model, equations are not used, but the use of statistics is implied, such as in the use of contingency tables needed to test relationships. Statistical analyses are necessary to implement the EM. But both Rubin’s

Causal Model and the Elaboration Model may be applied through a number of statistical methods; neither model necessitates a single type of statistical analysis.

Representation of Data

Because RCM and the EM consist of a numerical component, the way in which each represents data should be compared. Rubin's Causal Model, as exemplified in Table 2, requires that each individual participant be potentially exposed to all levels of the independent variable. Therefore, each participant occupies a row in Table 2, and contributes a number for each column. In the EM, each person belongs to a single combination of levels of the variables used, and has an outcome in only one condition; there is no comparison among conditions within persons, hypothetically or literally.

RCM, therefore, at its most basic level, involves hypothetical comparisons within individual persons. The EM involves testing data from groups and subgroups of participants. The VT applies to groups of participants (i.e., experimental group, control group), not to individual persons. The VT threats to validity, as well as its notation for designs, apply to an entire study, not the individuals involved in the study. The VT does not concern the nature of the data, although specific threats to statistical conclusion validity may differ in importance for different types of data (e.g., a greater sample size is needed when the data are of a categorical nature).

Application to Various Research Designs

These models differ with respect to their applications to various research designs. In RCM, the definition of a causal variable and the equations for estimating a causal effect are based on a (conceptual) true experiment in which an independent variable is manipulated and each participant is thought of as having been in the treatment and in the

control conditions. Holland & Rubin (1983) clearly state, though, that this model may be applied to nonrandomized as well as randomized studies. In their 1966 book, Campbell & Stanley specify which threats to validity may be particular to certain study designs. For example, when random assignment has been used in a study, selection is often not a plausible threat to validity. The VT may be applied to a study of any design. As discussed above, the Validity Typology does not apply to specific study designs - it applies to the inferences of a study.

The EM's utility seems to lie in quasi-experimental studies and surveys. When groups cannot be considered equated before a study takes place, or groups are created on the basis of measured information about the participants (e.g., if participants were classified by their responses to a survey item), any conclusions about the dependent variables measured may suffer from numerous alternative explanations. By choosing third variables to test the relationship between the original variables, the researcher searches for variables that could make the groups of participants interchangeable (Zeisel, 1968) (note: interchangeable does not mean that the participants are 'equal').

CHAPTER 8: CAUSAL LANGUAGE

Language communicates the conceptualization of relationships in a study.

Through language, a researcher chooses how to present theoretical background, hypotheses, methods, results and conclusions. Publishing research, particularly in the social sciences, requires careful attention to language choices because the readers, (students, researchers, policy makers), use this language to support or refute their own ideas.

Language denotes the type of relationships under study – whether they are associational relationships such as, ‘watching violent television programs *relates to* higher levels of aggression in children,’ or causal relationships such as, ‘watching violent television *causes* a higher level of aggression in children.’ Each of these two statements carries implications for television watching.

The degree to which plausible rival hypotheses (Cook & Campbell, 1979) can be ruled out impacts the perceived truthfulness, or validity, of a statement. The associational statement above conveys the sense that other variables must be considered before identifying the strength of the relationship between violent television and aggressive children. The causal statement above conveys the sense that this relationship is conclusive, violent television does cause children to be more aggressive. The causal statement could be used as evidence of the need to censor programs for young audiences, while the associational statement leaves some uncertainty about the direct role of television watching.

Causal statements differ from associational and predictive statements because a causal statement endorses a specific variable as causing an effect. An associational

relationship identifies the existence of a relationship between two (or more) variables (Holland, 1986a). A predictive relationship (i.e., a directional association), acknowledges that knowing the value of one variable helps us anticipate the value of another variable.

These types of relationships are not mutually exclusive. Certainly if a relationship is causal, this subsumes associational and predictive relationships between the causal variables. But a hierarchy of power, perhaps silently assigned to these types of relationships, still exists. Policy makers and expert witnesses may carry more sway when reporting causal conclusions that may be translated into clear support for a given program or policy (Rein & Winship, 1999). When policy makers form programs based on the goal of changing or improving an outcome, they are interested in manipulable variables (i.e., those variables that can be changed), and the subgroups of people for which this change is relevant.

Creating Scientific Knowledge

The discussion of how a researcher approaches writing a research article is an important one because research articles result in the creation and dissemination of scientific knowledge. The pieces of a study, the literature review, hypotheses, methods, measures, analyses, results and conclusions only form a coherent story when joined together through a sequence of ideas and logic.

For example, when a researcher decides to cite previous work, or builds upon the conclusions of another's work, the language used to explain the 'facts' of a study influences the reader's perception of the 'facts' and frames how the researcher builds a case for his/her own study. Bruno Latour writes about the creation of scientific knowledge, explaining the construction of a research article as a process in which the

researcher picks and chooses the works she wishes to cite to support her own ideas (Latour, 1987). Each statement depends on the statement that comes after it and so on (Latour, 1987) creating a chain of ideas that join to communicate the story.

The writing of a scientific research article entails more than mere fact reporting (Sternberg, 1988). In Sternberg's popular reference book, The psychologist's companion: A guide to scientific writing for students and researchers, he advises, "Successful scientific writing must both inform and persuade. Students often believe that a successful piece of scientific writing need only inform the reader of the scientists' data and their interpretation of the data. The reader is then left to decide whether the theory provides a plausible account of these (and possible other) data. This conception of scientific writing is incorrect" (1988, p. 8-9).

Before and during the creation of the paper stage, when variables are hypothesized to be causally related, is causality inherent in the relationship studied or is causality created by explanations imposed by the researcher? Although this dissertation does not focus on discourse analysis, the field of critical discourse analysis (e.g., Potter, 1997) would surely support the latter – that researchers impose a structure of causality in a research article. The concept of cause brings a sense of coherence to ideas (Gagnon, 1996). Discourse performs a function (Potter & Wetherall, 1987). One might consider causality as a function of discourse. In social science, this function, in part, is to make the story of a research article flow and to close, somewhat, the circle of possible explanations for a behavior or attitude. Sternberg's approach to writing the research article envelops the function of the causality idea; when a researcher shapes the final article to be

presented, the structure imposed on the article must offer some degree of continuity and explanation to the reader.

Interpreting Language

What words or phrases convey causal thinking? Nouns such as ‘effect,’ ‘consequences,’ and ‘influences’ relay the idea of cause, as do verbs such as ‘affect,’ ‘results in,’ ‘leads to’ and ‘produces.’ Table 8 provides examples of causal language and Table 9 provides examples of associative, predictive and ambiguous language. These examples provide only a sample of a wide array of words, nouns and verbs, used to communicate causal thinking. At times, researchers make pointed statements about cause and the idea of cause clearly comes across to the reader. Other times, cause may be indicated in a more subtle way, and close attention must be paid to the message being conveyed.

A researcher is socialized to some ‘rules’ of language, what should and should not be said. For example, in introductory research methods and statistics classes, students learn that correlation does not equal causation. Implicit in this rule is the idea that the language of association and prediction, not causal language, should be used in the interpretation of regression analysis. Students learn that causal language applies to experimental studies with random assignment of participants and a manipulable independent variable. After reading Cohen and Cohen’s (1983) seminal work, Applied multiple regression/correlation analysis for the behavioral sciences, one may become somewhat confused about this rule because Cohen and Cohen freely use causal language to describe and interpret regression analysis.

But researchers often do use causal language when reporting regression results. A common practice for reporting regression includes language such as “Variable X accounted for Y% of the variance in variable Z” and “the direction of the effect was...”. These standard phrases for reporting regression results include causal language but may not be used in accordance with the research design choices and theoretical considerations that would warrant causal language to be used in the reporting of statistical results.

This standard of reporting raises the question of whether or not researchers intend or interpret this standard language as causal in relation to regression. If, when reading such phrases, researchers recognize the language as causal but consciously choose not to interpret it causally because of regression-reporting conventions, this standard way of reporting may be in need of reconsideration. What consensus, if any, do researchers agree upon for the interpretation of regression results when regression is used for prediction versus causation?

The fifth edition of the American Psychological Association’s Publication Manual (APA, 2001) devotes 38 pages to “Expressing ideas and reducing bias in language.” The issue of causal language receives no attention in the manual. Even when distinguishing between the use of ‘since’ and ‘because,’ APA only suggests, “*Since* is more precise when it is used to refer only to time (to mean “after that”); otherwise, replace with *because*” (p. 57). APA does advise, when discussing precision and clarity in writing a research article, “Make certain that every word means exactly what you intend it to mean. Sooner or later most authors discover a discrepancy between the meaning they attribute to a term and its dictionary definition. In informal style, for example, *feel* broadly

substitutes for *think* or *believe*, but in scientific style such latitude is not acceptable” (p. 36-7)

Perhaps it is assumed that researchers will carefully select words to describe the type of relationships between variables – causal, associational, predictive. Certainly this is not meant to suggest researchers only use causal language in the context of randomized experiments. With the existence of thorough methods to assess the validity of causal inferences (e.g. Validity Typology, Elaboration Model, Rubin’s Causal Model), certainly myriad research designs and analyses, quantitative or qualitative, may be used to test causal hypotheses. But given the importance of using precise and clear language in results and conclusions, especially for the audience who will use the results and conclusions of an article, researchers should carefully choose their words.

CHAPTER 9: METHOD

Rubin's Causal Model, the Validity Typology, and the Elaboration Model offer different approaches to assessing causal inferences. For decades, social scientists have recognized the Validity Typology and the Elaboration Model as enhancing the accuracy with which causal conclusions are interpreted. To evaluate whether or not, and how, research can further benefit from the application of Rubin's Causal Model, all three models were applied to published research studies and the resultant interpretations compared to assess the unique and overlapping contributions of each model.

Testing models and methods by using invented and extant examples (Maxwell & Cole, 1995) and through examining published research (Lazarsfeld, 1955) are common techniques utilized by scientists who research methodological issues. The American Psychological Association's description for its journal Psychological Methods reads, "Where appropriate, submissions should illustrate through concrete example how the procedures described or developed can enhance the quality of psychological research." In this study, applications of the three models to the chosen examples lead to the creation of a list of recommendations for researchers to follow when evaluating causal inferences.

Data

Research articles qualified as the most appropriate data to use for this study because the issue of causality can best be addressed in the context of the written research article. Once the researcher completes a study and reaches the stage of preparing the research manuscript, s/he must choose how to present the research. Choices in presenting research include framing the research issue using selected theory and previous research

findings, wording hypotheses, operationalizing variables, choosing data analysis strategies, interpreting results and drawing inferences.

For this analysis, three published research articles were analyzed according to each methodological model (RCM, VT, EM). Applying these models requires specific knowledge from each selected article: hypotheses, specific information about independent and dependent variables used, the method of data analysis, as well as investigating the match between statistical analysis and the researcher's interpretation and presentation of the conclusions.

The article selection process proceeded by reviewing articles published in relevant journals in the past five years, (a span of five years was chosen so that the methodological issues in the articles cannot be considered out of date) and selecting ones that fit the following criteria, presenting possibly interesting challenges to Rubin's Causal Model. Articles chosen as data for this study passed the following primary selection criteria: 1) causal language had to be used in the article's hypotheses, results or conclusion section, 2) no random assignment of participants to conditions was desired; however if an article did involve random assignment and some aspect of the research posed a challenge to Rubin's Model it could still be selected for analysis, and 3) all articles came from journals published by the American Psychological Association (APA).

Causal language must be used in the hypotheses, results or conclusion of the article to ensure that the researcher was working from a causal frame of analysis. Language serves as a primary tool for classifying the types of relationships a researcher pursues. Nonrandomized as opposed to randomized studies were preferred because it is in situations where randomization has not taken place that the greatest challenges arise for

inferring valid causal effects. However, given the definition of a causal variable provided by Rubin's Causal Model, studies using a noncausal variable as a manipulated variable were also considered, regardless of randomization, because such studies present additional challenges to the model.

Articles were required to be published by APA because this professional organization is regarded as one of the most respected sources of research in the social sciences. Choosing articles published by this organization decreases the chance that the causal inference issues investigated are a matter of poor research quality. Articles published by the American Educational Research Association were also considered for selection. In the end, in order to choose articles from different disciplines, the three selected articles came from the education, personality and social psychology, and industrial/organizational journals from APA.

Once articles fit the above three criteria, the following more specific criteria were applied to choose the final articles. Articles were required to include variables presenting a challenge to the tenets of RCM, such as demographic variables and personality variables, which RCM considers attributes, not causal variables. A range of research designs were also desired so as to contribute an array of methodological issues to examine through the models.

No systematic way of choosing these articles was implemented because the goal was not to collect a representative sample of articles from a field, or to generalize to a population of studies. The goal was to choose articles that entail prototypical problematic methodological issues encountered in research, (as identified by classic texts such as Cook & Campbell, 1979 and articles such as Lord, 1967, as well as by general knowledge

of the field of psychological research), and would therefore present useful settings for comparing the utilities of these models.

In total, three research articles were chosen for analysis: 1) “Parenting and academic socialization as they relate to school readiness: The roles of ethnicity and family income” published in a 2001 issue of the Journal of Educational Psychology, 2) “Is all sexual harassment viewed the same? Mock juror decisions in same- and cross-gender cases,” published in a 2001 issue of the Journal of Applied Psychology, and 3) “Who attains social status? Effects of personality and physical attractiveness in social groups,” published in a 2001 issue of the Journal of Personality and Social Psychology.

These three articles, all published by the American Psychological Association, differed in several respects. Articles 1 and 3 included naturally occurring groups, while article 2 involves random assignment. In article 2, with random assignment, the variable manipulated by the researcher was biological sex, which Rubin’s Model would not classify as causal. All three articles contain causal language (obvious as well as not-so-obvious) in several sections throughout.

Each article includes demographic or personality variables, and approaches the study of these variables in a causal way. Article 1 analyzed the influence of ethnicity, while article 2 manipulated the biological sex of persons through vignettes and photographs, and article 3 investigated the effects of personality variables and physical attractiveness. Article 1 contained data gathered through surveys, article 2 measured participant’s responses to controlled situations, and article 3 combined self-reports, peer-reports, and independent observations. Articles 1 and 3 mainly used correlational and regression analyses, while article 2 used analysis of variance.

Procedure

In order to collect the information necessary for applying each model, each research article was first summarized. A summary included a brief overview of the article, and a discussion about and examples of the language used to describe relationships in the study. After summarizing each article, each of the three models (RCM, VT and EM) was applied to every article. Each of the three articles was analyzed separately, and the applications of the models are presented in succession for each article. A summary of results from the three models is presented after the analysis of each article. Then, comparisons among the models and articles occur in the discussion chapter(s) following presentation of all results.

Below are general guidelines followed when applying each model to assess the validity of causal inferences from a study. Not all steps applied to every study, and some articles presented challenges to these applications. With this in mind the steps below form general procedures to follow when using these models.

Applying Rubin's Causal Model

Applying Rubin's Model may lead to the following possibilities: 1) the model, its notation and equations will help specify the types of relationships studied and assumptions involved in those relationships, 2) the notation and equations will not fit well with the nature of the variables or hypotheses (e.g., they may prove too simplistic or too specific to particular types of studies). Either or both of these possibilities further the goal of assessing what RCM can offer social science research. To begin the analysis of RCM, I followed the model's notation. This notation has been used, for example, by Holland &

Rubin (1983) in applications of this model, by Kaufman & Cooper, (1999), and by Wainer (1991).

The following procedure was implemented:

1) Classify each variable according to RCM framework (Holland & Rubin, 1983):

- a) S: Does the independent variable fit the definition of a causal variable? If yes:
 - Determines to which condition a person is exposed
 - What are the possible levels (e.g., t and c)?
- b) G: If the variable does not fit the definition of a causal variable, is it an attribute?
 - A sub-classification variable
 - What are possible values (eg., male, female)?
- c) Y: What is the outcome variable (s)?
 - What are the potential values of Y?
 - Outcome, has as many versions as there are levels of S
 - Eg., if S can equal t or c, Y_t and Y_c would be the values of Y
- d) X: Are there any covariates (these may be continuous and causal)?
 - An individual should have the same value for X no matter what S equals. If the variable was affected by treatment, it is a concomitant variable, not a covariate, and should not be controlled for to obtain a causal effect estimate.

2) Through the process of classifying variables in step 1, the independent variable (s) will be classified as causal or not causal (S vs. G variables). Given the nature of these variables, what types of conclusions did the researcher make?

- If the researcher drew causal conclusions, then are the necessary pieces present to be able to apply RCM equations for a causal effect? (e.g., is there a control condition present?)
 - If the necessary pieces are not there, what assumptions are being made to reach the causal conclusion?
 - How plausible are these assumptions?
- If not causal conclusions, then what kind of conclusions can be drawn from the data?
- Were attribute variables used? In what capacity (e.g., independent variable, covariate)?
 - If attribute variables were used, were they considered as proxies? What types of conclusions were drawn with respect to the attribute variables?

Applying the Elaboration Model

The following steps can be used as a guide to implementing the Elaboration Model. As stated above, these steps may be refined as unanticipated situations arise in the scenarios/studies being analyzed. Although these steps may seem basic, their implementation is often less straightforward when in the context of the intricacies of a particular study.

- 1) Did the researchers establish a bivariate relationship between the hypothesized 'cause' and 'effect' ?
- 2) If so, did the researchers consider any additional variables (i.e., test variables) to elaborate on the original bivariate relationship? Did they consider under what conditions the bivariate relationship may or may not exist?

- 3) If the researcher's intent was only to investigate a bivariate relationship, are there other variables that could have been included in the analysis to elaborate this relationship? Were any of these variables measured?
- 4) If no relationship was found between the two original variables, did the researchers still proceed with further analyzing the nonsignificant relationship? For example, searching for a suppressor or distorter variable?
- 5) When trying to understand causal order or how to articulate the theory involved, where the test variable falls in the chain becomes key. Is the test variable intervening, antecedent or neither?
- 6) Taking into consideration where the test variable falls in the chain, what type of partial relationships were created once the test variable entered the analyses? Table 7 presented earlier in this paper will be used as a guide for classifying partial relationships.
- 7) After carrying out these steps, the study's conclusions will be considered. Did the conclusions drawn from the analyses match the results from the analyses? That is, if the conclusion includes a causal inference, is this inference warranted from the specific variables analyzed and the analyses conducted?
- 8) Because the use of the Elaboration Model in the literature concerns mainly contingency tables and categorical variables, in the process of applying this model to the scenarios and studies chosen for this paper, possible challenges may arise when other types of statistical analyses have been used. It is difficult to know ahead of time what these difficulties may be; it is usually while testing out the model when such issues become apparent.

Applying the Validity Typology

The research design of a scenario/study is first described using the notation of the VT as illustrated earlier in the paper. This highlights whether or not participants were randomly assigned to groups, and presents a timeline for when measurements and treatment implementations took place in the study.

Then threats to validity are considered. For each of the four types of validity, there exists a list of possible threats. To apply this model, the plausibility of each in these lists of threats is evaluated in relation to each study. Depending on the design and elements of the study, plausible threats are either confirmed or disconfirmed and explained.

Some threats are more relevant than others to questions of causal inference. Specifically, threats to internal validity pose the greatest problems for establishing causal conclusions because these threats offer possible alternative explanations for why a cause-effect relationship may exist. Some threats to construct validity, as well, pose challenges to inferring a valid causal inference. Those threats that describe potential influences on participant behaviors during the course of a study, or that influence the way in which the treatment is measured or experienced, generate circumstances that can obscure the relationship between variables. In the list below, the most relevant threats to causal inference are marked with an asterisk. These are the threats that will be primarily addressed in the analysis.

Threats to Validity (Shadish et al, 2002)

Internal validity.

- * **Ambiguous temporal precedence:** Is there confusion about the temporal order of variables (i.e., uncertainty regarding which variable is the cause and which is the effect)?
- * **Selection:** Could the way participants were allocated into conditions depend on a variable that could be biasing the results? If so, any differences found may be due to these preexisting difference between groups, and not the treatment.
- * **History:** Were there other events (besides the treatment of interest) that may have influenced participants during the course of the study?
- * **Maturation:** Were there psychological and/or biological changes that may have naturally occurred in the participants during the course of the study that could have influenced the outcome?
- * **Regression:** Were participants selected on the basis of extreme scores? If so, it is possible that those participants who were chosen for their high scores will not score as high on a subsequent measurement, and those participants chosen for their low scores will score higher on a subsequent measure, even in the context of no treatment (particularly if the measure is unreliable).
- * **Attrition:** Is the participant drop-out rate comparable among the study groups? Or is it different, possibly due to differences between those participants who stayed versus left the study? If more participants drop out of one particular group, this could bias the results, creating a selection problem. There could be characteristics of those participants who remained in the study that may contribute to the treatment effect found.

- * **Testing:** Could there have been effects of having taken a test on measurements collected during subsequent testing? Sometimes participants' responses change as they become more familiar or bored with a measure or measurement technique.
- * **Instrumentation:** Does the reliability (stability) of the person or instrument measuring the dependent variable change during the study? If the measurement instrument changed as the outcomes were assessed, the resulting treatment effect includes the variance of this measurement.
- * **Additive and interactive effects of threats to internal validity:** "The impact of a threat can be added to that of another threat or may depend on the level of another threat" (p. 55).

Statistical conclusion validity.

Low statistical power: Could the results from the study incorrectly demonstrate no relationship between the independent and dependent variable (e.g., was the sample size too small, were covariates taken into consideration, were cell sizes too unequal)?

Violated assumptions of statistical tests: Did the data conform to assumptions of tests used?

Fishing and the error rate problem: Were too many tests conducted without taking the significance level into account?

Unreliability of measures: Are the measures accurately assessing the variables?

Restriction of range: Are there constraints on the possible values of the variable, or are the levels of a variable too disparate for meaningful comparison?

Unreliability of treatment implementation: Are all participants in a condition receiving the same level or concentration of the independent variable?

Extraneous variance in experimental setting: Are there distracting conditions in the study environment that may affect the relationship between the independent and dependent variables?

Heterogeneity of units: Is there a great deal of variability among participants within a study group? Are there variables, on which the participants differ, that are related to the outcome, but not the treatment? If so, this will inflate the error variance.

Inaccurate effect size estimation: Were the correct analytic procedures used (e.g., were procedures for continuous variables used only for continuous variables?) and possible statistical problems eliminated (e.g., outliers)

Construct validity.

Inadequate explication of constructs: Does the label of a construct match the way in which it is operationalized? E.g., Is the measurement used too specific to capture the entirety of the construct? Is it too general to accurately assess the construct under study?

* **Construct confounding:** Is more than one construct being captured by the way in which a variable is operationalized?

Mono-operational bias: Is only one instance of a treatment or one assessment of an outcome made?

Mono-method bias: Was only one method used to implement treatments or collect data? If so, is the analysis one of the data or a combination of the data and the method used?

Confounding constructs with levels of constructs: Do the conclusions drawn represent the specific levels of the construct used?

- * Treatment sensitive factorial structure: Did exposure to a treatment possibly differentially change the way in which groups understood a measure, leading to differences in responses that could be hidden in scoring procedures?
- * Reactive self-report changes: Could participants have initially provided inaccurate information to receive certain treatments, and later responded to outcomes in a more accurate way, clouding the actual treatment effect?
- * Reactivity to the experimental situation: Does the treatment construct take into account that the way in which participants experience the study setting could impact their responses?
- * Experimenter expectancies: Were participant responses influenced by indications from the experimenter about how to respond? If so, the treatment effect may reflect the experimenter's influence and will not be an accurate assessment of the treatment.
- * Novelty and disruption effects: Could the newness or radical nature of a treatment have contributed to the treatment's effect? If so, this must be taken into account for the "treatment construct description." (This same rule holds for confirmation of the following three threats.)
- * Compensatory equalization: Did the researcher provide benefits of some kind to participants not receiving treatment?
- * Compensatory rivalry: Could participants have performed better than usual because they wanted to show they could do as well as those receiving more beneficial treatment?
- * Resentful demoralization: Could participants have performed unusually poorly because they were upset about not receiving the benefits available to other conditions?

* Treatment diffusion: Did participants cross over to receiving treatment from a condition they were not assigned to?

External validity.

Interaction of the causal relationship with units: Does the causal effect hold for different samples (i.e., samples with varying attributes)?

Interaction of the causal relationship over treatment variations: Is the causal effect found specific to the levels or kinds of treatment used?

Interaction of the causal relationship with outcomes: Does the causal effect apply only to the outcome variable used in the study?

Interactions of the causal relationship with settings: Is the effect found specific to the place and environment in which the study was carried out?

Context-dependent mediation: A significant mediating variable may not be relevant in situations other than the one used in the study.

Applying the Regression Framework

Most quantitative social science studies include data analyses using regression or a type of analysis based on regression. When a study's results have been analyzed using this framework, the way in which the analysis was carried out will be examined.

Specifically:

- 1) Was the analysis correctly set up to answer the hypotheses posed?
- 2) Were the hypothesized relationships the actual relationships tested?
- 3) If control variables were used, how were they selected? As discussed earlier in the paper when the regression framework was explained, specific issues should be considered when selecting a control variable. Of particular interest to this paper is

the way in which Rubin's Causal Model would approach the issue of selecting controls: Given that attribute variables cannot be causal in that model, does this present a problem for using them as controls?

- 4) Do the ways in which the data were interpreted match the results from the analyses?

Interpreting the Data

Each article chosen for analysis presented unique challenges to Rubin's Causal Model. After implementing each model for every article, and articulating any complications arising from their applications, the discussion section expands upon the places where RCM offers useful considerations as well as places where social science offers useful considerations to the structure of the model.

The discussion section presents a methodological/philosophical reflection of the perspective that each model takes on a study. By comparing the conclusions that *were* drawn versus conclusions that *would be* drawn according to the models, several areas of confusion and clarification receive attention. Then, based on the analysis, a list of recommendations is presented for researchers to consider when planning research or interpreting and applying results.

CHAPTER 10: ANALYSIS OF ARTICLE 1:

HILL (2001), “PARENTING AND ACADEMIC SOCIALIZATION AS THEY RELATE TO SCHOOL READINESS: THE ROLES OF ETHNICITY AND FAMILY INCOME”

This first article selected for analysis was published in a 2001 issue of the Journal of Educational Psychology. This article fits the selection criteria in that it includes causal language in the introduction, hypotheses and conclusions, is a nonrandomized study, and is published by the American Psychological Association. In addition, as desired in studies chosen for this analysis, a key variable in this article, ‘ethnicity,’ falls under Rubin’s categorization of attribute variables. The study reported in the article is a cross-sectional survey, using regression analyses to analyze the data. Mothers and schoolteachers provided the data for predictor variables; children provided the criterion data.

The study aimed to identify predictors and explanations of children’s school readiness as measured through two scales: “The Sound-Letter Correspondence scale” to measure pre-reading level and “The Quantitative Concepts scale” to measure pre-math skills. The predictive and explanatory variables included parenting style, parental involvement in school, parental expectations for the child’s educational success, ethnicity and income. In this chapter, all quotes from Article 1 are accompanied by only a page number, with the understanding that this is the article under analysis, and the author and date are already known from the chapter title.

Causal Language

Identifying causal language in the article entailed considering the obvious, as well as searching for not-so-obvious, causal language. Obvious causal language includes

verbs such as “affects,” “influences,” “causes,” and nouns such as “effect.” Obvious associative and predictive language was identified through verbs such as “predicts,” “is associated with,” “relates to.” Less obvious verbs indicative of a cause-effect relationship include “creates,” “results in,” “leads to,” “buffers,” “moderates.”

In this article, the researcher used both associative and causal language when reporting results of previous research in the introduction. The types of relationships hypothesized for the actual study were causal, as exemplified through the use of the noun “impact” and the verb “influences” in the following hypotheses quoted from the article, with my added underlining to highlight causal language: “Thus, in the present study, I examined the relationships between parenting behaviors and parental academic socialization and children’s school performance among young children. I also looked at the impact of various types of parental behaviors, expectations, and parental involvement in school, I was particularly concerned with whether these relationships were generalizable between African American and Euro-American families and among different family income levels. Specifically, I studied the separate and interactive influences of family income and ethnicity on children’s school readiness at kindergarten by comparing samples of African American and Euro-American children and their families that are comparable on socioeconomic indicators” (p.687).

The results of the regression analyses, as presented in the ‘Results’ section, were communicated through associative and predictive language. Language such as “Variable X explained x% of the variance in Y” and “the direction of the effect was....,” was not considered purposeful causal language because it is standard in reporting of regression results.

Verbs such as “were associated,” “were related,” and “correlated with” communicate the relationships resulting from the statistical analyses in this study. The possible exception to this associative language is the verb “moderated,” which is used throughout the paper to discuss the relationships of ethnicity and income with parenting variables. The relationship between an IV and DV depends on the level of the moderator variable, i.e., there is an interaction of the IV and the moderator. Therefore, moderation could be considered a causal concept. The use of the term “moderate” in this article indicates some level of causal thinking about ethnicity and income. Baron and Kenny’s (1986) seminal article about moderation and mediation presents moderation as a causal, not associational or predictive, concept.

In the discussion/conclusion section, where elaboration of the regression results and applications of those results takes place, causal language became more prevalent. The discussion section begins with a mix of causal and associational language. The restating of regression results without statistical references (e.g., no numbers) was mostly done using associative language. But then the follow-up sentences explaining implications of the results or restating them outside the context of the study (i.e., generalizing them) were often causal. The distinction seems to be that in the discussion/conclusion section, the farther the purpose of a statement was from reporting a statistical result, the more likely it seemed to be a causal statement.

There was some inconsistency in the way results were reported in the discussion section when describing ethnicity as moderating the relationship between specific parenting variables and premath score: “For African-Americans, parents’ involvement at school enhanced children’s premath performance” (p. 694). This is a causal statement

about the effect of parental involvement for African American parents. For African Americans, parental involvement at home and teacher perceptions about parental education values were not related to premath scores. “Therefore, encouraging African American parents to involve themselves in school may improve their children’s math performance” (p. 694). “For Euro-Americans, parents’ involvement with educational activities at home and teachers’ perceptions about the extent to which parents valued education were positively related to premath performance” (p. 694).

Note the use of causal language to describe one interaction result for African Americans, and associational language describing the result for Euro-Americans. Although the language in this conclusion is inconsistent, there does not seem to be an obvious purposeful reason why one statement was made to be causal and the other not causal. However, after studying the introduction, hypotheses, and conclusions, it was established that the researcher was conceptualizing the relationships in this study in a causal way.

Applying Rubin’s Causal Model

S: Exposure Variable

The variable ‘S’ indicates the condition to which a participant is exposed. In the literature on RCM, examples usually refer to levels of ‘S’ as the ‘treatment’ condition or the ‘control’ condition of a study. In this study, there is no manipulation; no variables are manipulated, there are no treatment and control conditions. The researcher’s hypotheses, and the RCM definition of a causal variable, must be used to determine which variables involved in the study are ‘S’ variables.

According to the researcher's hypotheses with my added underlining and italicized words in brackets, "Thus, in the present study, I examined the relationships between parenting behaviors and parental academic socialization and children's school performance among young children. I also looked at the impact [*causal*] of various types of parental behaviors, expectations, and parental involvement in school, I was particularly concerned with whether these relationships were generalizable [*indicating subgroupings*] between African American and Euro-American families and among different family income levels. Specifically, I studied the separate and interactive influences [*causal*] of family income and ethnicity on children's school readiness at kindergarten by comparing samples of African American and Euro-American children and their families that are comparable on socioeconomic indicators" (p.687).

As indicated by the word 'causal' in parentheses after each variable hypothesized to have a causal relationship with the outcome 'school readiness' (as determined by the use of 'impact' and 'influences'), all of the variables tested in the regressions in this study are hypothesized to be causal variables. How do these variables fare with the RCM causal definition, that in order for a variable to be causal, it must be manipulable and the process of the unit being exposed to various values of the causal variable should not change anything else about the unit under study?

To determine this, it must be decided whether or not the following variables fit RCM's definition of a causal variable. Are the variables 'parental behaviors', 'expectations' and 'involvement at school' manipulable? Are 'family income' and 'ethnicity' manipulable?

Parenting Behaviors

To the extent that parental behaviors, expectations and involvement are not personality variables (e.g., shy parents may not want to get involved in school activities), these are potentially manipulable variables. But they are complicated potentially manipulable variables, unlike a treatment versus control condition.

For example, parental behaviors may be parent, as well as child, specific. The way parents treat their children, whether parents act hostile and controlling or accepting, may have a lot to do with parents' social interactions in general, or with the way in which their parents raised them. What is the potential for manipulating parental behaviors? It would be possible to ask a parent to interact with their child differently, but that style of interaction may not work with the child, and may feel unnatural for the parent. But to apply Rubin's Model, the question does not involve whether or not parents could act differently, whether parental behavior is manipulable in and of itself. The relevant question is whether or not the *child could be exposed to different parenting practices*.

The important point to consider for the application of Rubin's Model is the possibility of the unit under study being *potentially exposed* to different levels of parenting behavior, parental involvement, and parental expectations. Could a child be potentially exposed to a different level of parenting behavior? Is it possible to imagine that a child would be exposed to a parent who is warm and accepting or be exposed to a parent who is hostile and controlling? If we imagine the situation in which a child is adopted, this potential exposure is clearly possible. If we imagine a scenario in which a child's parent divorces and remarries, this child could be exposed to different parenting

behaviors. Given the potential for a child to be exposed to different levels of parenting, parenting behavior qualifies as a causal variable.

In this study, parental involvement in home, as well as in school, was measured. Home involvement was measured by asking moms about their involvement; school involvement was measured by asking moms and teachers to complete different scale items and then combining the scores from these two sources. It does seem possible that a child could be exposed to different levels of parental involvement, and the level of parental involvement a child is exposed to could be manipulated. This may be difficult in reality with the multiple responsibilities of parents (e.g., work), but is certainly possible to imagine. Given the possible collinearity between these two worlds, disentangling the influence of home and school involvement on a child's school readiness might be difficult.

'Parental expectations' becomes a bit more difficult to imagine in terms of a causal variable under Rubin's Model. Could a child potentially be exposed to different parental expectations? Where do parental expectations come from? Here, the variable of time becomes important because perhaps parental expectations are an effect rather than a cause of child school readiness. If a child is not performing well, her parent may develop low expectations for her. Or, to think about this relationship in the opposite direction, a child may not perform well because her parent's expectations of her are low.

This discussion of how to classify parenting variables raises an important distinction to keep in mind when applying Rubin's Model. One could easily become confused, in this study, as to whether the manipulability of parenting variables should be considered from the point of view of the parent or the point of view of the child. For

example, the concern when classifying parenting behaviors as causal or not causal is not whether we can manipulate the behaviors of parents. The concern is whether or not a child could be exposed to a different level of parenting behavior, that is, manipulating the child's exposure to parenting. This is a fine distinction arising precisely because the values of the independent variables originate in one group of units (i.e., parents), and the values of the dependent variables originate in a separate group of units (i.e., children). One can become misled when not approaching the issue of manipulability from the perspective of the unit supplying the outcome measures.

According to Rubin's classification of types of studies, this study would fall into the 'uncontrolled causal study' category. It is 'uncontrolled' because the independent variables are not under the researcher's control. It is a causal study because the researchers intend to investigate causal questions and are using some causal variables (i.e., parenting variables).

Parenting practices, parental involvement, and parental expectations fit RCM's definition of a causal variable. For each of these variables, a child could potentially be exposed to a different value. This is not to say that if a child was exposed to different parenting practices or expectations, that other things about the child would not change. It is possible to imagine a child exposed to different values of these variables without the child drastically changing in the process. However, over long periods of time, one could imagine that aspects of a child's attitudes or behaviors may change according to new parenting practices, involvement, and expectations.

Income

Returning to the question of whether or not ethnicity and income are causal, we can imagine a child's family income being different than what it is. But in reality, when someone's income changes, this effects other critical things that relate to income, such as education, material possession, privileges like access to health care, and a person's point of view [e.g., standpoint theory (e.g., Collins, 1991) – where you stand in society relates to how you understand social issues]. Would all these things have to change if, in a study, a person's income were to be changed from \$20,000/year to \$40,000/year? No. Not in the moment that the change would take place, and not necessarily in a moment soon after, either. Therefore, income qualifies as a causal variable according to RCM.

The hypotheses presented income as both a causal and as a subgrouping variable. Subgrouping variables are 'G' variables, and 'G' variables are not causal in Rubin's Model. When income was used as a subgrouping variable, conclusions drawn would be descriptive, according to RCM. A causal variable may also be used descriptively (i.e., if two variables causally relate, then they are associated), but the reverse is not true (i.e., if two variables are associated, they are not necessarily causally related).

G: Subgrouping Variable

What qualifies a variable as a G variable? Holland & Rubin (1983) report that a G variable is one for which the value could not have been different for a person. In this paper, the hypotheses pose ethnicity and income as subgrouping variables *and* as causal variables: "Thus, in the present study, I examined the relationships between parenting behaviors and parental academic socialization and children's school performance among young children. I also looked at the impact [*causal*], of various types of parental

behaviors [*causal*], expectations and parental involvement in school, I was particularly concerned with whether these relationships were generalizable [*indicating subgroupings*] between African American and Euro-American families and among different family income levels. Specifically, I studied the separate and interactive influences [*causal*] of family income and ethnicity on children's school readiness [*effect*] at kindergarten by comparing samples of African American and Euro-American children and their families that are comparable on socioeconomic indicators" (p.687).

According to RCM no causal conclusions can be drawn about ethnicity. Ethnicity is not considered a causal variable in RCM. It is considered an attribute, a subgroup variable, which the model labels with the notation 'G.' A person's ethnicity could not change without other things about that person changing. But this also depends on how one operationalizes and defines ethnicity. In this study about parenting, ethnicity is operationalized as the family either being African American or Euro-American. Categorization based on self-report. Although this is not a causal variable in RCM, the author of the research study does develop causal hypotheses, as discussed above, about ethnicity. RCM would classify ethnicity as a 'G' variable – a subgroup variable.

Y: The Outcome Variable

It is not an easy task to figure out the levels of Y, the outcome variable, in this study. Examples provided in the literature on RCM include categorical independent variables, not continuous independent variables. In this research article, the three parenting variables, parenting practices, involvement and expectations, are measured through continuous scales. For example, the variable 'parenting practices' consists of 5

subscales: acceptance, enforcement, withdrawal of relations, hostile control and inconsistent discipline. Each item of the subscales is rated on a 0-4 scale.

When a continuous scale is used as a causal variable, how is the equation for a causal effect constructed? Subscales are used as the variables entered into the regression analyses, so the causal effect equation should also be constructed at the subscale level. But when there are multiple levels, such as in the case of a continuous variable, how is it decided which levels will be compared to which others, and how comparable the difference will be between various levels?

For example, a child could potentially be exposed to a 0, 1, 2, 3, or 4 on the acceptance subscale. The causal effect for acceptance is the child's reading score when parents express X level of acceptance minus child reading score when parents express Z level of acceptance:

$$Y_x - Y_z$$

But what values are appropriate to use for X and Z? If a child's parent reports an acceptance level of 2, to what score should this value of 2 be compared in order to calculate the causal effect equation? The causal question here is how the *degree* of acceptance, not the presence versus absence of acceptance, affects child's school readiness. This example emphasizes that a causal effect results from the *comparison* of conditions or scores.

Parenting variables such as maternal acceptance are subscales, with each item rated 0-4 on a Likert scale. Given that the other subscales were not included in a single regression equation, this is not a comparison of maternal acceptance with enforcement or hostile control, for example. Because maternal acceptance was the only parenting

variable in the regression equation, the researchers examined the degree to which maternal acceptance existed and how this degree related to reading and math scores. For example: $E(Y | S = a)$, where a equals level of acceptance and can take any value from 0 to 4.

A result about acceptance was: “Maternal acceptance was positively associated with prereading and premath scores.” One would obtain the causal effect of acceptance by comparing the expected values of children’s school readiness at various levels of acceptance. This analysis requires knowing the functional form of the relationship. The functional form refers to the regression line formed when relating the parent score of acceptance with the child’s outcome score. Once the *shape* of this line becomes known, the researcher possesses the necessary information to construct meaningful comparisons of levels of acceptance.

For income the causal effect equations are much easier to fill in because it is a categorical variable. Income, by nature, takes a continuous form, but here it was used as a dichotomous variable. For the hypotheses posing this variable as causal, S would be the relevant tool to develop the causal effect equations: S would have two levels: low and not low. The causal effect equation would read: $E(Y_{low} - Y_{notlow})$

When the variable ethnicity is posed as a subgrouping variable, which would be a G variable in RCM, the causal effect equation would read:

$$E(Y | G = ea) - E(Y | G = aa)$$

In this equation, Y can only take on only a single value for each participant, and the results can only be interpreted descriptively.

Validity of Causal Inferences

After applying the notation of the model, the next step is to consider the validity of the conclusions drawn in the article. Through the process of classifying variables as done above, the independent variable (s) were classified as causal or not causal. Given the nature of these variables, what types of conclusions did the researcher infer and what would RCM say about these conclusions?

From the above discussion, RCM would not identify problems with using parenting variables as causal variables. However, as discussed above in the section in ‘Y’ variables, the application of the causal effect equations for these variables may present problems, leading to difficulty in evaluating the validity of causal conclusions. When the equation consists of a comparison between two conditions (eg., treatment and control), one easily constructs the causal effect equation by inserting values for responses to these conditions. Applying the equation to continuous independent variables necessitates more information than comparisons involving two separate conditions such as a treatment and control.

In the case of a continuous independent variable, applying the causal effect equation requires knowing the functional form of the relationship between the continuous variable and the outcome. This functional form may be linear or non-linear. The form’s shape does not matter for causal analysis.

To aid in the application of the causal effect equations to continuous independent variables, the model may require extended notation to indicate that the independent variable can take any value in a given range, and that the comparison between any two values in this range produces a causal effect estimate. For example, while $E(Y|S = t) -$

$E(Y|S=c)$ communicates the causal effect value for those exposed to the treatment and control conditions of a study (i.e., a study with a categorical independent variable), $E(Y|S = t_1) - E(Y|S = t_2)$, where t can take on any value between 0 and 100, for example, communicates the causal effect value for participants exposed to any two levels of a continuous variable. In an equation for a continuous independent variable, there is no control condition; the conditions of comparison are various levels of the treatment.

Basic information about the functional form is reflected in the reporting of conclusions. For example, “Mothers’ expectations for grades were positively related to children’s prereading and premath performance” (p. 694) indicates that the higher the score on expectations, the higher child readiness scores were.

The hypotheses included income as both a causal and subgrouping variable. Both associational and causal conclusions about income appear. Conclusions about income did not suffer validity problems, but they did demonstrate inconsistencies: “Although many of the relationships observed in this study are generalizable across ethnicity and family income levels, family income moderated [*causal*] the relationships between some parenting behaviors and prereading performance. The relationships were stronger [*associational*] for lower income families than for higher income families. Parenting practices and the parent-child relationship are important influences [*causal*] on children’s school performance in low-income contexts. Thus, the stress associated with many low-income environments that often diminishes [*indirectly causal*] early school performance may be buffered [*causal*] by supportive parenting [*note that here, parenting is positioned as the moderating variable*].” In this last sentence, the researcher seems to reposition the

variables, placing income as the main causal variable, and parenting as a moderating variable.

The conclusion continues with variables resulting from income status as causing school readiness, placing income, indirectly, in a causal role, “Among children from higher income families, often there are fewer obstacles to interfere with (*indirectly causal*) school performance and a greater number of resources to enhance (*indirectly causal*) performance or make up for (*indirectly causal*) poorer parenting practices (citations). Hence, parenting behaviors may buffer (*causal*) some of the disadvantages associated with (*associational*) low income status. This is consistent with McLoyd (1990) who proposed that poverty influences (*causal*) child outcomes through its effect (*causal*) on economic stress and parenting practices” (694).

Again, positioning ‘parenting’ as the variable that can interact with, or buffer, income to produce school performance reverses the original hypothesis that income would moderate the relationship between parenting and school readiness. In addition, reasons are offered for why the relationship between income and school readiness exists. The last sentence of this quote clearly communicates income causing child outcomes, and positions parenting as a mediating variable, through which income acts on child outcomes.

Conclusions drawn about ethnicity were difficult to classify as causal versus not causal. Although no direct phrases could be identified as saying that ‘ethnicity caused,’ careful reading of the conclusions suggests the researchers may have been extra cautious about communicating the idea that *because* a parent is African American, she may not

use as good a parenting style or be as involved in her child's school life as a Euro-American mother.

Conclusions convey that ethnicity moderated the relationship between parenting variables and school readiness: "Perhaps because Euro-Americans are more involved at school than their African American counterparts they have had more opportunities to learn how to assist their children at home in math-related tasks" (p. 694). Mainly the discussion includes the fact that 'ethnicity moderates' but is that not a valid statement? The validity of this statement and others like it depends on the interpretation of moderation.

What can we conclude of the type of relationship studied through moderation? Are these causal relationships? Baron & Kenny (1986), a commonly referenced article about mediation and moderation, describe a moderating variable, "In general terms, a moderator is a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and or/strength of the relation between an independent or predictor variable and a dependent or criterion variable" (p. 1174). Using a clearly causal verb, 'affects,' it seems that discussing ethnicity as a moderating variable places it in a position to be considered causal variable. Throughout Baron & Kenny's (1986) article, they use causal language to explain mediation and moderation.

In the following paragraph of results about ethnicity, ethnicity is discussed as a causal variable. Notice there is no direct causal language identifying ethnicity as a cause of school outcomes, but the verb "moderated" indicates ethnicity as a causal variable. Also notice that there is no direct associative or predictive language explaining the link between ethnicity and school outcomes. Instead, the implications are that Euro-American

parents are more involved in school than African American parents and therefore their children display different outcomes. The suggestion that African American parents need encouragement to participate in their children's schooling indirectly suggests that ethnicity is a cause, whether proximal or distal, of school readiness. In the quote below, I have placed words in italicized parentheses to indicate causal and associational language used to describe relationships.

“Ethnicity moderated [*causal*] the relationship between parents' school involvement, including the extent to which parents value education, their involvement at school, and their involvement at home, and premath performance. For African Americans, parents' involvement at school enhances [*causal*] children's premath performance. However, involvement at home and teachers' perspectives on the extent to which parents valued education were not significant. Therefore, encouraging [*causal*] African American parents to involve themselves in school may improve [*causal*] their children's math performance. For Euro-Americans, parents' involvement with educational activities at home and teachers' perceptions about the extent to which parents valued education were positively related [*associational*] to premath performance. Perhaps because [*causal*] Euro-Americans are more involved at school than their African American counterparts (citations) they have had more opportunities to learn how to assist their children at home in math-related tasks” (p. 694)

The last sentence accompanying the above quotation reads, “However the reading readiness of all children may benefit from encouraging positive relationships between parents and teachers.” (694-5), thereby generalizing a need for improving parent-teacher relations regardless of income or ethnicity.

Rubin's Causal Model would not support any causal conclusions about ethnicity as valid causal conclusions. Holland (1986a) would not support these as valid because a child in this study could not potentially be exposed to another ethnicity without other aspects of this child changing. Rosenbaum & Rubin (1984) do not support causal conclusions about aspects of a person that could not, in reality, be changed, and Rubin (1986) would not support these as valid conclusions either, because the 'treatment' ethnicity cannot be clearly defined or implemented for each individual.

Applying the Elaboration Model

According to the introduction and hypotheses of the research article, which guide our understanding of the researcher's intent, it seems that the primary relationships of study are parenting behaviors and school readiness. Next, is an interest in how these relationships generalize to mothers/children of different ethnicities and income levels. There is also an additional interest in the effects of ethnicity and income on school readiness separate from parenting practices. The Elaboration Model was not designed to apply to the study of two main relationships, but will be examined here in the context of the first relationship between parenting and school readiness, with ethnicity and income added as test variables.

The first step in the Elaboration Model is the establishment of a relationship between the main independent and dependent variables of interest. In this study, correlations were run to assess relationships among the independent, dependent and moderator variables. That is, correlations were tested for relationships between parenting, ethnicity, income and school readiness.

According to the correlation table presented in the paper, 7 of the 15 parenting independent variables were related to reading ability and 6 of the 15 related to math ability. Regression analyses were tested for all parenting variables. Even though relationships were not found for over 50% of the independent variables with the dependent variables, a test variable was still added into the analysis. Ethnicity was not related to reading ability, but was related to math ability, and income was related to both abilities.

Having analyzed bivariate relationships through correlations, regression analyses were then run to assess the individual and interactive relationship of ethnicity and income on school readiness. Finally, income and ethnicity were added into regression equations with parenting variables to determine their role as moderators.

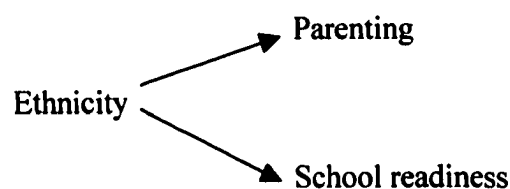
The general logic of these analyses resembles the logic of the elaboration model, first testing bivariate relationships, then adding test variables to examine how the original relationships changes. However, the Elaboration Model does not completely match the order in which variables were entered into these regression analyses. This issue will be touched upon here in relation to how the analyses would be conducted following the EM's logic, and will be further discussed in the 'regression framework' section below.

According to the EM, the regression analyses testing the moderating role of ethnicity and income would first include a test of the parenting variables with school readiness. Then, ethnicity or income would be entered and the resulting relationships between parenting and school readiness examined.

Examples of the Elaboration Model in the literature typically include only categorical variables; in this study parenting variables and school readiness variables are

continuous and ethnicity and income are categorical. Given the continuous IVs, contingency tables cannot be used to test the partial relationships. With continuous variables, test variables can be tested through regression, for example, by adding another variable into the regression equation to control for a test variable, or by adding an interaction between two or more variables.

As time ordering of the test variables is an important component of the EM for classifying the types of resulting partial relationships, where do ethnicity and income fall in the order of relationships? According to the Elaboration Model, ethnicity and income precede parenting variables and school readiness because values of ethnicity can be determined from the beginning of one's life, as can income. In addition, parenting preferences may be guided by ethnic customs and traditions, and a parent's income may guide the amount of time she is able to spend at home or school. Therefore these variables fall under the category of 'antecedent' variables because they would have occurred before parenting and school readiness variables. It is important to note that categorizing these as antecedent variables creates a time ordering that places parent variables as well as school readiness as dependent variables to ethnicity and income.



The point should also be made here, as was raised in an earlier section about control variables and order of entry in Chapter 6, that the occurrence of variable values in

our actual lives may not present the most useful gauge for determining the ordering of variables in a theoretical framework. When the effect of a variable occurs in a theory may be a better gauge of order (D.M. Rindskopf, personal communication, December 18, 2002).

When a test variable occurs antecedent, three types of partial relationships may result: replication, explanation and specification. Interpretation no longer remains an option because in order for a variable to be considered as interpreting a relationship between two other variables, it must intervene in the time order sequence of these other two variables.

Taking into consideration where the test variable falls in the chain, what types of partial relationships were created once ethnicity and income entered the analyses? Table 7 presented earlier in this paper will be used as a guide for classifying partial relationships.

Resulting partial relationships could be considered as falling into two categories:

- 1) When the addition of ethnicity and income into the regression analysis (here these variables were added in the form of an interaction with parenting variables) was not followed by a change in the relationship between parenting and school readiness, the partial relationships would be replication, because the original bivariate relationship did not change, or did not change much, with the addition of these variables.
- 2) When the interaction was significant, the partial relationships fell under the 'specification' category because the additional of the third variable specified the relationship between the original two variables.

An example of a test variable resulting in replication is the following: “Maternal acceptance was positively associated with prereading and premath scores. Hostile socialization strategies were negatively associated with prereading and premath scores” (p. 690). These two sentences demonstrate significant bivariate relationships between a predictor and an outcome.

The next sentence, “The interaction terms of ethnicity with each parenting variable, added at Step 3, were not significant and did not significantly add to the variance explained” (p. 690) demonstrates that the addition of ethnicity as a test variable did not change or explain the relationship between the original predictor and outcome. This is not to say that if ethnicity had been added as a control variable the original bivariate relationship might not have changed; that is still a possibility, but because ethnicity was entered before parenting variables in the regression analysis, that cannot be determined from the analysis. The lack of an interaction in this instance cannot completely rule out the role of ethnicity. But in the analyses conducted here, ethnicity entered the analysis in the form of an interaction and its presence did not change the original bivariate relationship.

The following sentence exemplifies ‘specification’ as a result of adding ‘income’ into the regression. This sentence involves the same predictors as the above paragraph exemplifying replication: “However, the addition of the interactions between each parenting variable and family income added significantly to the amount of variance explained for sound letter-correspondence score....For the relationship between each parenting variable and sound-letter correspondence scores, the slope was steeper for the lower income families than for the higher income families” (p. 690). This sentence

demonstrates that ‘income’ specifies a condition under which these parenting variables predict the outcome. Adding the variable ‘income’ specified that the relationship between parenting and reading ability was stronger for lower income parents than for higher income families.

In the context of the Elaboration Model, these results are problematic though. The order in which variables were tested in the regression analyses does not follow the logic of the EM or general rules for testing moderation in regression. From the analyses conducted, it is not clear how the addition of a test variable changed the original bivariate relationship because the bivariate relationships discussed in the results are bivariate relationships controlling for ethnicity and income, i.e., partial correlations. Therefore, the bivariate relationships between parenting and the outcomes discussed as results of the regression analyses should, but do not match the bivariate correlations between parenting and the outcomes.

Applying the Validity Typology

This study included neither a treatment nor a control group; multiple groups of people (moms, teachers, kids) were measured at a single time point, and then the relationships among variables from these groups were assessed. Therefore the independent variables of interest were measured simultaneously with the dependent variables, with the assumption that the IVs occurred in time before the DVs. Shadish, Cook & Campbell (2002) classify this design as ‘nonexperimental,’ indicating the correlational nature of the study and the lack of random assignment, pretests or control groups.

‘Parenting practices, involvement and expectations’ were considered the main independent variables, providing an ‘X’ assumed prior in time to the outcome measures of readiness, which are represented by ‘O.’ Given that all measurements were taken at one point in time, no subscript number exists for this ‘O’. Groups used in this study include preexisting groups, that is, groups existing prior to the study. These groups as defined by the study were European-American and African-American mothers and children and low-income and high-income mothers and children. With parenting as the main independent variable, preexisting groups were used, no random assignment took place, and these groups were not exposed to different treatments, so they are not represented by different X’s.

The following notation represents the study design:

	X	O
NR	-----	
		O

Where ‘X’ represents ‘good’ parenting practices and the omission of X represents not good parenting practices. The application of this notation feels like somewhat of a stretch, though, because the VT emphasizes manipulability and describes the use of ‘X’ as a treatment variable. The model does not provide a separate notation for independent variables that cannot be manipulated, notation to distinguish between manipulable and nonmanipulable independent variables, or a specific notation to represent continuous independent variables.

Internal Validity

Internal validity is of serious concern in this study. Immediately, *ambiguous temporal precedence* becomes an issue: Is there confusion about the temporal order of variables (i.e., uncertainty regarding which variable is the cause and which is the effect)? The researchers acknowledge this issue explaining that they do not know the direction of relation for parental expectations, for example, whether expectations start with the parents or is formed on impressions of the child.

The threat of '*selection*' probes the way in which participants ended up in groups in a study, questioning whether or not other variables related to group formation may be biasing the results. In this study, children were already selected into homes with a particular parenting style. Groups were not formed by the researcher. African-Americans and Euro-Americans were selected for the study because of their ethnicity, and their extant income levels. In this situation, several other variables related to ethnicity, such as discrimination and stereotypes could have played into teacher reports, as well as parental expectations. The income a mother reported could have been related to the number of hours she must be employed in order to provide for her and her child, thereby decreasing the amount of time she could possibly be involved in her child's schooling.

In the 'Introduction' section of the paper, the researchers acknowledge a host of variables related to ethnicity and income that may contribute to specific parenting practices: "Because family income is related to parental education level, lower income parents may have had less positive experiences with the schools and thereby be less able to provide their children with the skills to negotiate the school system successfully, which may negatively affect school performance (citations)" (p. 687).

Several potential differences were investigated, such as “There were no ethnic differences in the number of children in the home,” a variable that could possibly contribute to the time a parent has to be involved in the child’s schooling. However, “Although there were no differences in children’s exposure to schooling before entering kindergarten, African American children were more likely to attend Head Start than were Euro-American children” (p. 688). In the statistical analysis though, nothing was done to account for this difference. The list of alternative explanations is quite extensive.

Attrition did not seem to be a significant problem in the study. But during the recruitment and measurement stages, some participants did not continue with participation. Nine mothers who agreed to be contacted and were eligible for the study did not follow through, and “All teachers agreed to participate, however several teachers did not return their forms after several reminders” (p.689). No further discussion ensued about the teachers or parents who did not continue with the study.

Instrumentation does not seem like a significant threat to the study. Four trained interviewers conducted the interviews to collect data. There was probably no more variance within one interviewer’s data collection than there was across interviewers. All interviewers had experience “working with ethnically and socioeconomically diverse children and families,” and received training on the interview protocol.

Internal validity threats not relevant to the current study include *history, maturation, testing, additive and interactive effects of threats to internal validity.*

Construct Validity

Some construct validity threats also present points of consideration in this study.

Construct confounding presented a significant threat to this study. Construct confounding asks: Is more than one construct being captured by the way in which a variable is operationalized? The relevancy of this threat pertains to the variables ‘ethnicity’ and ‘income.’ As explained by the researchers in the ‘Introduction’ section, many variables important for parenting and school readiness are related to ethnicity and income. Important to the idea of construct validity is that many of these variables may be captured by a measure of ethnicity or income.

The only mention of confounding in the ‘Introduction’ came in the discussion of the interrelation between ethnicity and income: “Such differences based on income and ethnicity are often difficult to interpret because family income is often confounded with ethnicity” (p.687). But the list of variables captured under the guise of ethnicity and income would be a cumbersome one to untangle.

Mono-method bias concerns the lack of multiple methods used to implement treatments or collect data. If only one method is used, this method may contribute to the findings. For children, it seemed that the interviewers administered instructions for the scales the child would complete, and the child completed the scale alone. It seemed that mothers were responding verbally to scale items administered by the interviewer, as well as completing scales on their own. It is unclear how teachers’ responses were measured. Even though all data was collected through scales, some of these seemed verbally administered, while others were self-administered.

Turning to the issue of *reactivity to the experimental situation*: Does the treatment construct take into account that the way in which participants experience the study setting could impact their responses? Social desirability bias could be acting in a study like this,

for parents who may not admit that they treat their children poorly, or have low expectations of them.

Threats to construct validity important for causal thinking but not relevant for this study include *mono-operational bias, confounding constructs with levels of constructs, treatment sensitive factorial structure, reactive self-report changes, experimenter expectancies, novelty and disruption effects, compensatory equalization, compensatory rivalry, and resentful demoralization.*

Applying the Regression Framework

The researchers chose regression as the statistical analysis for this paper. Most of the independent variables were continuous (all parenting variables) and some were categorical (ethnicity and income). Categorical variables coded as dummy variables fit into the regression framework. Because only two values of ethnicity were studied, African-American and Euro-American, only one dummy variable is used; this variable would be coded as 0 to represent one level of ethnicity and 1 to represent the other level. This same logic applies to 'income' because income also consists of two levels. Both outcomes in the analyses were continuous.

Two sets of regression analyses were run. The first included ethnicity, income and the interaction between ethnicity and income as predictors. The second included parenting as a predictor and the interaction between ethnicity and parenting or between income and parenting, with ethnicity and income tested as moderating variables.

Typically, tests of moderation include variables entered in the following order as exemplified in equation 22 below: First, the main IV of interest (X1) enters, then the

hypothesized moderator (X_2), then the interaction between the IV and moderator (X_1X_2).

$$Y = B_1X_1 + B_2X_2 + B_3X_1X_2 + e \quad (22)$$

In this analysis of parenting, the order in which variables were entered into the regression analysis does not follow the typical order recommended for a moderation test. “After entering ethnicity as Step 1, the addition of the parenting variables at Step 2 explained significantly more variance... The interaction terms of ethnicity with each parenting variables, added at Step 3, were not significant” (p. 690). Here, the moderating variable was entered before the main IV, meaning that any interpretation of the relationship between parenting and the outcome from this regression is controlled for ethnicity. This would lead to the relationship between parenting and the outcome seeming different than the bivariate correlation between the two indicators.

As far as interpreting the interaction, this ordering of the first two variables does not make a difference. The amount of additional variance explained by the interaction does not depend on the order of entry of the first two variables because at that point, the interaction is tested controlling for both ethnicity and parenting. But ethnicity was added as one would enter a control variable into a regression, which it is not in the case of a moderating relationship.

Additionally, Baron & Kenny recommend (1986) “...it is desirable that the moderator variable be uncorrelated with both the predictor and the criterion (the dependent variable) to provide a clearly interpretable interaction term” (p. 1174). In

these analyses ethnicity was only related to 2 of the 15 parenting variables and related to math ability but not reading ability. Income related to 4 of the 15 parenting variable and both outcome variables.

Summary of Results After Applying Each Model to Article 1

Applying each of the three models, RCM, EM and VT, highlighted drastically different issues in this article about parenting, income, ethnicity and school readiness. These applications also highlighted important areas of possible expansion in each of the three models.

According to the definitions of Rubin's Causal Model, the validity of causal conclusions about parenting variables and income suffer no problems, but interpreting ethnicity in a causal manner does present a problem. In this article, ethnicity plays a moderating role. Moderation communicates a causal relationship: The relationship between an independent and dependent variable depends on the level of the moderating variable. Here, the relationship between parenting and school readiness depends on the ethnicity and income of the family.

Direct causal language appears in conclusions about the moderating role of income, but this presents no problems given that income qualifies as causal for RCM. As far as interpreting moderation in general (i.e., outside the scope of article 1), can one interpret it causally sometimes and descriptively other times depending on the nature of the moderating variable? This would seem inconsistent, as the causal interpretation should lay with the process of moderation and not the moderating variable. Looking to leaders in the field of moderation and mediation, Baron & Kenny (1986), one finds both concepts explained as clearly causal.

According to the Elaboration Model, the validity of causal conclusions in this article does not suffer any threats. However, inconsistencies in the way in which relationships were tested become apparent when applying the EM framework to this article. The logic of the EM presents a clear method of testing relationships, beginning with bivariate relations and then elaborating with test variables. In the context of article one, this logic did not apply so clearly because the order in which variables entered the analysis seemed inconsistent.

According to the Validity Typology, threats to internal validity existed. More than one possibility existed for the temporal sequence of some variables, for example, parental expectations as relating to child's abilities. Selection of participants presented a threat because children had been selected into certain parenting practices, involvement and expectation, and income and ethnicity groups existed prior to the study. The conclusions also suffered threats to construct validity, such as the confounding of constructs for ethnicity and income and the possibility of parents reacting to the research situation (e.g., social desirability).

Weaknesses/areas for growth that the article highlighted in the models included:

1) additional notation needed for continuous independent variables in RCM, 2) the distinction of applying the potential exposability idea to the unit of study and not the manipulable independent variable in RCM, 3) the needed specification of statistical models to test each type of partial relationship in the EM, 4) the difficulty in applying the VT notation to the continuous, uncontrolled independent variables (i.e., parenting variables). These issues will be discussed in detail in chapter 13 of this dissertation.

CHAPTER 11: ANALYSIS OF ARTICLE 2:

WAYNE, RIORDAN & THOMAS (2001), "IS ALL SEXUAL HARASSMENT VIEWED THE SAME? MOCK JUROR DECISIONS IN SAME- AND CROSS-GENDER CASES"

The second article selected for this analysis was published in a 2001 issue of the Journal of Applied Psychology from the American Psychological Association. This article fits the selection criteria in the inclusion of causal language, the use of random assignment with manipulation of biological sex, (a variable challenging Rubin's Model), and the use of attribute variables: gender of participants, gender of persons in vignettes. Although the term 'gender' is used throughout the article, it refers to the biological sex of persons in vignettes and the participants themselves. Because the researchers used the term 'gender,' I will use their term throughout the discussion of this article. In this chapter, all quotes from Article 2 are accompanied by only a page number with the understanding that this is the article under analysis and the author and date are already known from the chapter title.

Participants in this study were randomly assigned to one of four experimental conditions. To create the conditions, the researchers designed vignettes about a sexual harassment case, and manipulated the gender of the sexual harasser (male or female) and target of sexual harassment (male or female), creating four conditions from all combinations (same gender and cross gender) of harasser and target gender. Participants read one of the four vignettes and saw pictures of the harasser and target; participants acted as mock jurors in this study, reading one of the four vignettes and looking at a photo of the harasser and a photo of the target.

The study investigated the relationship between gender of a sexual harasser, gender of the target of harassment and gender of the participant playing the role of mock juror assessing the harassment case. Gender of harasser, target and participant were the explanatory variables. Participants provided the outcomes by judging the verdict for harasser, monetary award for victim, and reporting their perceptions of the case.

Causal Language

Causal language is abundant in the introduction section of this article when referring to previous research as well as when referring to the current research study: “The purpose of the present study was to investigate three understudied gender categories of SH (female harasser-male target, male harasser-male target, female harasser-female target) plus the traditional category (male harasser-female target) to determine whether decisions in a SH case differ based on the gender of those involved” (p. 179).

The hypotheses, however, do not include language that is obviously causal. For example, “In cross-gender cases, mock jurors will (a) assign more guilty verdicts, (b) assign greater monetary awards, and (c) perceive the behavior as more serious, offensive, and inappropriate when the harasser is a man (with a female coworker target) than when the harasser is a woman (with a male coworker target).” There are no causal words in this hypothesis, but the implication is that the gender of the harasser and target will *lead* the jurors to provide certain ratings as opposed to other ratings. All the hypotheses in this article are of this nature – not directly causal, but with causal implications.

There is no causal language in the results section when presenting statistical analyses, beyond the restatement of the hypotheses, which involve causal implications. Causal language becomes prevalent again in the discussion section. As is written in the

first paragraph of the discussion section, with my added underlining to highlight causal phrases “Overall, we found there were several combinations of harasser and target gender that resulted in more guilty verdicts, monetary awards, and perceptions that the behavior was more inappropriate, serious, and offensive than other gender combinations, although the behaviors described in the cases were identical” (p. 184).

All results from this study were formed into causal conclusions – whether they involved the effects of the ‘manipulated’ variables or the participant gender: “These findings suggest that although there are many other factors that influence SH decisions, the gender composition of the harasser, target, and participant does play a role in harassment decisions” (p. 186). “...we still found that participant gender had some effect” (p. 186).

Applying Rubin’s Causal Model

S: Exposure Variable

Holland and Rubin define S as “... a variable that indicates the experimental condition to which each unit in P is exposed” (1983, p. 6) (P being the population of units.) In this study, there are two variables manipulated by the researcher: harasser gender (male or female) and target gender (male or female). Harasser gender and target gender are both ‘S’ variables. The gender of the harasser and target are manipulated to create four conditions, two of which are cross-gender and two of which are same-gender. These four conditions are manipulated via two operationalizations of gender, vignette wordings and photographs. The vignette and photograph genders matched in each condition; each participant was exposed to only one harasser gender and one target gender.

The two cross-gender conditions are 1) a male harassing a female (from here on indicated by the subscript 'mf') and 2) a female harassing a male (indicated by the subscript 'fm'); the two same-gender conditions are 1) a male harassing a male (subscript 'mm') and 2) a female harassing a female (subscript 'ff'). Each participant could potentially be exposed to any one of these four conditions. Because there are four conditions to which each participant can potentially be exposed, S has 4 levels. In Rubin's symbols, the four levels of 'S' are as follows, with a subscript 'm' representing 'male,' 'f' representing 'female' and the first subscript representing the sex of harasser while the second subscript represents the sex of target:

$$S_{mf} \quad S_{fm} \quad S_{mm} \quad S_{ff}$$

To summarize, there are two 'S' variables, gender of harasser and gender of target. Each variable has two levels (male or female harasser and male or female victim) the combinations of which create four levels of exposure (four study conditions), and four levels of S.

G: Subgrouping Variable

"Unlike S, it is not possible for the value of G for each unit to have been other than what it is" (Holland and Rubin, 1983, p. 6). One variable RCM would clearly classify as a 'G' variable in this study is participant gender because this is an attribute of participants, it is not manipulated nor could it be. The variable 'participant gender' classifies participants into two groups: male and female participants: $G = m$ or $G = f$.

Notice that gender has now been classified as *both an 'S' variable and a 'G' variable*. How does this one variable classify as both variable types? Here, the issue of Rubin's definition of a causal variable and the causal effect equation enter as key

determining factors. The idea of potential exposability as the focal aspect of the causal variable definition excludes attribute variables from being classified as 'S' variables. Gender is an attribute variable. However manipulating gender in the vignettes demonstrates that when gender acts as an attribute of a person in the vignette as opposed to an attribute of the participant, it qualifies as a causal variable because it becomes potentially manipulable. The discussion chapter of this dissertation expands upon this seemingly contradictory classification.

Y: The Outcome Variable

Each level of 'S' has a corresponding version of 'Y' (i.e., how a participant would respond after exposure to a level of S). Here we have four levels of 'S', and their four corresponding versions of 'Y'. Three outcome variables were measured for each participant: 1) Verdict, 2) Award, and 3) Perceptions. Because details of the particular outcome variables are irrelevant for this analysis, I will proceed as if there is just one outcome variable.

Because four levels of 'S' exist in this study, there are four versions of 'Y.' How do we take the third independent variable 'participant gender' into account when considering possible versions of the outcome?

Each of the three independent variables has two conditions: harasser (male and female), target (male and female), and participant (male and female). Although the experiment has four conditions, the statistical analysis has:

2 (male, female harasser) $\times$ 2 (male, female target) $\times$ 2 (male, female participant) = 8 conditions, because the 2-level variable 'participant gender' is also studied. In the end,

Rubin's logic for this analysis would be understood as two levels of the 'G' variable for each of the four 'S' conditions to which participants could potentially be exposed:

$S_{mf} G = f$	male harasser, female target, (for a female participant)
$S_{mf} G = m$	male harasser, female target, (for a male participant)
$S_{fm} G = f$	female harasser, male target, (for a female participant)
$S_{fm} G = m$	female harasser, male target, (for a male participant)
$S_{mm} G = f$	male harasser, male target, (for a female participant)
$S_{mm} G = m$	male harasser, male target, (for a male participant)
$S_{ff} G = f$	female harasser, female target, (for a female participant)
$S_{ff} G = m$	female harasser, female target, (for a male participant)

Although the distinction between 'S' and 'G' variables does not matter in the usual way one thinks about an ANOVA, this distinction does matter for Rubin's model and subsequently becomes important for the interpretation of results from the typical ANOVA.

Four hypotheses were posed in this study. Some of these hypotheses investigate responses to the vignettes of all participants together (males and females) and others investigate differences between the male and female participants. Following are RCM's causal effect equations pertaining to each hypothesis.

Hypothesis 1: "In cross-gender cases, mock jurors will (a) assign more guilty verdicts, (b) assign greater monetary awards, and (c) perceive the behavior as more

serious, offensive, and inappropriate when the harasser is a man (with a female coworker target) than when the harasser is a woman (with a male coworker target).”

Using RCM notation, the causal effect equation for mock juror response in cross gender cases when the harasser is a man versus woman is:

$$Y_{mf} > Y_{fm}$$

This hypothesis is causal: that the combination of harasser and target gender causes differences in participant’s verdict, award, and perception. No levels of ‘G’ are specified in this equation because the hypothesis includes both male and female jurors together. As the elements of the causal equation are accounted for through this hypothesis, it is a causal hypothesis according to the tenets of Rubin’s causal effect equation. Comparing the outcomes from the ‘male harasser, female target’ condition and the ‘female harasser, male target’ conditions results in the causal effect of one condition relative to another. Participants could have been exposed to either of these conditions, and the independent variable of interest in the hypothesis was a manipulable variable.

But what can be said of the fact that the independent variables being manipulated in this hypothesis are gender variables? If Rubin’s Model does not consider attribute variables as causal (e.g. gender), but the causal effect equations can be constructed in such a way that a valid causal comparison takes place between levels of the attribute variable (gender), where does that leave the classification of conclusions about gender?

Hypothesis 2: “In cross-gender cases, female mock jurors will (a) find the harasser guilty more often, (b) assign greater monetary awards, and (c) perceive the behaviors as more serious, offensive, and inappropriate than will male mock jurors.”

Forming Rubin’s equation for this hypothesis:

$$[(Y_{mf} + Y_{fm}) | G = \text{female}] > [(Y_{mf} + Y_{fm}) | G = \text{male}]$$

Researchers posed this as a causal hypothesis, but in Rubin's Model, hypothesis 2 cannot be causal. The gist of this hypothesis focuses on juror gender as a cause, but participants cannot be exposed to another value of their own gender. Juror gender is an attribute of the participants. From the point of view of Rubin's Model, this qualifies as a descriptive hypothesis. The researchers treat hypothesis two the same as other hypotheses in this paper, but Rubin's Model does not.

Hypothesis 3: "Mock jurors will (a) assign more guilty verdicts, (b) assign greater monetary awards, and (c) perceive the behavior as more serious, offensive, and inappropriate when the harasser and target are of the same gender than when they are of the opposite gender."

$$Y_{\text{samegender}} > Y_{\text{oppositegender}} = (Y_{mm} + Y_{ff}) > (Y_{mf} + Y_{fm})$$

Hypothesis 3 compares same-gender harassment with opposite-gender harassment. Participants could be potentially exposed to either a same or opposite sex harassment condition. Therefore, Hypothesis 3 is causal. Holland (1986) reminds that a cause is always relative to another cause. Here, one condition is the cross-gender harassment, the other is same-gender harassment, and the causal effect is the difference between the two. This hypothesis presents the pieces necessary to compare for an effect.

Hypothesis 4: "In same-gender cases, mock jurors will (a) assign more guilty verdicts, (b) assign greater monetary awards, and (c) perceive the behavior as more serious, offensive, and inappropriate when the harasser and target are both men than when the harasser and target are both women."

$$Y_{mm} > Y_{ff}$$

This hypothesis is similar to hypothesis one, two conditions are being compared, both the gender of the harasser and the gender of the target change from one condition to the other. Given that the elements of the causal effect equation are satisfied and participants could potentially be exposed to either condition, the effects of interest in this hypothesis qualify as a causal effect.

X: Covariate

X represents a covariate. The researchers did test previous experience with sexual harassment and social desirability as potential covariates but these variables demonstrated no significant relationship with the outcomes, so were not included in the formal analyses.

Validity of Conclusions

Through the process of classifying variables as ‘S’ or ‘G’ the independent variables have been classified as manipulable or not. Given the nature of these variables, what types of conclusions did the researcher infer and what would RCM say about these conclusions?

The conclusion section summarizing the results reads as follows, with my added underlining to highlight causal phrases: “Yet the results of this study suggest that the gender composition of the harasser and target is an extralegal factor that exerts a moderate influence on verdicts (accounting for 8% of the variance) and perceptions (accounting for 14% of the variance; Cohen, 1977). Participant gender exerts a statistically yet small (1%) influence on SH ratings. These findings suggest that although there are many other factors that influence SH decisions, the gender composition of the

harasser, target, and participant does play a role in harassment decisions” (186). Clearly the results corresponding to all hypotheses were interpreted in a causal manner.

What would RCM say about these causal conclusions? The model clearly classifies gender as an attribute, and therefore noncausal, variable. But as the applications of the causal equations demonstrated, this study about gender includes the necessary pieces to complete the equations for hypotheses about differences between conditions. The causal equations can be satisfied, but the variable defining the conditions that comprise these equations is a noncausal one.

Why does this contradiction arise? Because gender of person in the vignettes defies Holland’s (1986) basic reason for attributes not being causal; that is, because exposure to different levels of the attribute includes other changes in the participant being exposed. But when the attribute belongs to a hypothetical person, and not the participant under study, Holland’s reasoning does not apply. Neither does Rubin’s reasoning, which was discussed in detail earlier in this paper, and will be revisited in the discussion chapter.

Rubin’s Model certainly would not support the inference of any causal conclusions about juror gender, such as the statement in the above quote about participant gender having an “influence on SH ratings” and the following statement: “Contrary to propositions by (citation) that participant gender differences would disappear if [*possibly causal*] more rigorous scenarios were used, we still found that participant gender had some effect [*causal*]” (186).

Rubin’s Model, as originally developed, would venture to say that this is a descriptive study because gender is an attribute variable, and therefore no causal

conclusions may be inferred from its use. This has interesting implications because the authors want these results to be considered in terms of managerial decisions and juror decisions in sexual harassment matters. “If evidence from this study generalizes to the courtroom or the boardroom, it suggests when a female employee brings a hostile work environment claim against a male coworker, jurors or supervisors may be less likely to find him guilty than if the situation was reversed.” Surely Rubin’s Model would problematize the identification of gender as a causal agent in an actual sexual harassment case in which the gender of an individual person is causally related to his or her guilt. How would Rubin’s Model explain to a juror or to a manager that the biological sex of the harasser or victim is not the reason why sexual harassment may have occurred?

Proxy Variables

Were gender variables considered as proxies? Perhaps to some degree they were in the conclusions, but not in the design and measurement stages. For example, the following conclusion substitutes ‘sex roles stereotypes about the target’s receptiveness’ in place of the variable ‘gender’: “Interestingly, the present finding suggests that sex role stereotypes about the target’s receptiveness to the behavior may not adequately explain SH decision making” (184).

The researchers attempted to explain the gender results with reasons why people may have rated the male/female harasser differently. The discussion section contains several statements about why particular results occurred, but not so much in terms of the idea that gender itself might not be the cause of differential ratings on the part of the participants: “More theoretical and empirical attention is clearly needed to understand the underlying processes used when evaluating the gender context of SH” (184).

The discussion section does include *explanations* of why the results may have occurred, but this differs from the concept of a proxy variable. The proxy variable concept directly acknowledges that a variable may be used to represent other concepts highly related to it but not actually measured. When considering for what variables biological sex could act as a proxy, several come to mind in the context of sexual harassment, such as power differences, gender norms, and physical strength to resist harassment, to name a few.

Applying the Elaboration Model

This study cannot easily be explained in terms of the Elaboration Model. The EM's focus on a bivariate relationship elaborated upon through the addition of a test variable precludes its applicability to a study, such as the present one, in which the original bivariate relationship is difficult to identify, and multiple test variables are added.

In the introduction section of article two, the need to study combinations of harasser and target gender beyond the male harasser-female target combination was acknowledged: "The purpose of the present study was to investigate three understudied gender categories of SH (female harasser-male target, male harasser-male target, female harasser-female target) plus the traditional category (male harasser-female target) to determine whether decisions in a SH case differ based on the gender of those involved" (p. 179). Some previous research probed differing perceptions between male and female harassers, but most previous research probed the relationship between male and female participant reaction to a male harasser, female victim.

The first step in the Elaboration Model paradigm is establishing a relationship between the hypothesized main cause and effect. Then, after exploring this bivariate

relationship, the next step considers under what conditions the relationship does or does not exist/change. Although no bivariate relationships were tested in this study of harassment, the study is built upon previous research of bivariate relationships.

Statistical results of bivariate relationships do appear in the 3-way ANOVA table presented in the paper, but there was never a discussion of the relationship between, for example, gender of harasser and an outcome, or gender of target and an outcome in this particular study. This is because all hypotheses concerned *combinations* of harasser gender and victim gender and how jurors rated harasser and victim according to these combinations. In this study, the relationships sought concerned trivariate not bivariate relations.

Every hypothesis involved gender as the main variable or cause. Gender was operationalized in three different ways: 1) gender of harasser and gender of victim in a vignette, 2) gender of harasser and gender of target in photographs, and 3) self-report of juror gender. Three variables, gender of target, gender of harasser and gender combination of harasser and target, created the conditions presented to participants. These three variables present a situation in which when any two of these variables become known, the value of the third is automatically determined. For example, if the harasser is a man, and if the target is a woman, then the gender combination is a cross-gender combination (as opposed to same-gender). Likewise, if the target is a man and the gender combination is same-gender, then we automatically know the harasser was a man.

This attention to the mutually exclusive relationship of any two of these three variables (i.e., after we know the gender of harasser, we either need to know the gender of victim OR the type of gender combination – knowing the value of either of these last

two variables determines the value of the other) complicates the application of the EM to this article.

Although the study seems fitting for an application of the EM – with the first hypothesis for example focusing on the relationship between gender of harasser and verdict - and on first read, it even seems that the hypotheses follow the path of adding explanatory test variables, the study does not easily fit within the EM framework, as will be demonstrated below.

A main reason why an application of the Elaboration Model does not naturally fit with the logic and the way in which analyses were carried out in this article is that while the hypotheses concern gender combinations, the analysis of these hypotheses concern only the variables ‘gender of harasser’ and ‘gender of target’ and an interaction of these two. The variable ‘gender combination’ never appears as a variable unto itself in the analysis, but as the explanation below demonstrates, gender combination is the key variable in the analysis.

According to hypothesis one (letter a), the hypothesized relationship is between gender of harasser (cause) and verdict (effect) in cross-gender cases, that jurors will assign more guilt to a male harasser in a cross gender case than a female harasser in a cross-gender case. The issue of only cross-gender cases being involved complicates the typical usage of contingency tables. Because the hypothesis explicitly focuses on the gender of the harasser, that jurors will rate the male harasser more harshly than the female harasser, ‘harasser gender’ is the main independent variable in the table. The dependent variable in this table is verdict, guilty or not guilty (although note that verdict

began as a dichotomous variable but was turned into a continuous variable). This creates a table with 4 cells. Table 10 matches the set-up represented by this hypothesis.

Note, though, that this hypothesis concerns only the cross-gender combinations male, female and female, male, limiting the analysis to only cross-gender combinations, and entering gender combination as a variable into the analysis. The opposite of cross-gender is same gender, and same-gender is the other level of this third variable involved in the hypotheses, albeit by default; it is not included as a relevant level in hypothesis one. See Table 11. Only the left side of Table 11 becomes relevant for hypothesis 1. That is, the side under the 'Cross-gender' label relates to hypothesis 1. The side under 'Same-gender' is irrelevant for the answers sought in hypothesis one.

The comparison of Tables 10 and 11 demonstrate the inclusion of three variables in the analysis for this hypothesis, verdict, harasser gender, and gender combination. One may at first consider 'gender combination' as the test variable in the analysis. But recall that only one level of the variable 'gender combination' matters for hypothesis 1. Using only one level of a test variable defeats its purpose of inclusion in an analysis, which is to partial the original bivariate relationship into each level of the test variable. 'Gender combination' does not qualify as a test variable.

[Note: Because the remaining hypotheses entail similar issues to those covered for hypothesis 1, they will not receive specific attention.]

Classifying the relationships according to the partial relationship scheme becomes irrelevant for the most part because the classification scheme for partial relationships refers to the change in relationship between two variables when an additional variable enters the analysis. In the case of these hypotheses, with the exception of juror gender in

hypothesis 2, there are no actual tests of bivariate relationships. Therefore, there are no relationships to compare to the partial relationships, making the partial relationship classification moot. It would not make sense to compare the partial relationships to previous research reported in the introduction because the methodology for the study in this article was changed in response to possible methodological flaws in previous research. Comparisons within this study would be the most accurate way to assess partial relationships. Although bivariate relationships were reported in the Anova table in the paper, the bivariate relations bear no relevance to the hypotheses in the paper.

In addition to the variables tested in the hypotheses, the researchers tested “...previous experience with SH and social desirability, as potential covariates and found that they were not significantly related to the dependent variables” (p. 182). Given that this analysis of previous experience and social desirability showed no relationship with the outcomes, the researchers may have considered that their inclusion in the analysis would result in what the elaboration model calls ‘replication,’ which means the relationship between the main variables of interest would not change upon inclusion of a test variable.

The following conclusion appears in the results section: “Thus Hypothesis 1a was not supported because the direction of these effects is opposite to what was predicted.” This result may be an appropriate place to consider a distorter variable (Rosenberg, 1968). As discussed in chapter 4, the inclusion of a distorter variable changes the direction of a relationship. Perhaps if a distorter variable entered the analysis, the hypothesis may have been confirmed. Or, perhaps, the results found accurately reflect the relationship and there effect is, in fact, opposite of what was hypothesized.

Applying the Validity Typology

Four conditions existed in this study, and participants were randomly assigned to one of the four conditions. Using the notation presented in the Validity Typology, ‘X’ represents the treatment, which included the gender combination of harasser and target presented in a vignette and accompanying pictures of the harasser and target. There was no control group in this study; due to the nature of the variables and theories being tested, the concept of a control group does not make sense in this context. But the study design did include four different treatment conditions (X with subscripts 1-4 represent the various conditions). ‘O’ represents the measurement of outcomes, all outcomes were assessed at one time.

X₁ O

X₂ O

X₃ O

X₄ O

Internal Validity

Internal validity was quite strong for this study. In addition to mitigating *temporal ambiguity* by manipulating the introduction of treatment and subsequently measuring the outcomes, to insure the elimination of an order effect of harasser/target gender, “the order of the presentation of the photographs was balanced within each condition.” This means that any results found could not be attributed to the fact that participants met the harasser or target first.

Other internal threats not relevant for this study include *selection, testing, history, maturation, regression, attrition, instrumentation, additive and interactive effects of threats to internal validity*. No internal validity threats appear to present a significant problem for causal inference in this study.

Construct Validity

The construct validity of this study does suffer a threat, though, due mainly to *construct confounding*. Two issues pertaining to construct validity arise. One issue deals with the fact that while two variables were changing in most comparisons, the conclusions about such comparisons reflected a focus on just one of the changing variables. Another issue arises in the operationalization of gender through photographs. Were participants reacting solely to the gender of the person in the photos, or were they also reacting to other characteristics of the person in the photo?

Issue 1: In this study, jurors are asked to provide a verdict about the harasser, an award rating for the target and the perception of the harasser's behavior. The hypotheses involve comparisons between conditions that vary in two ways: gender of harasser and gender of target. None of these comparisons involves the manipulation of a single variable. Therefore, no clear statements can be made about the causal effect of gender of harasser on its own or gender of target on its own. This creates confusion on the part of the reader, not on the part of the study authors.

For example, if a comparison only included a male harasser, male victim versus a male harasser, female victim, one could conclude that any difference between these two conditions was due to the gender of the victim. The same clarity results from comparing mm and fm conditions, ff and fm conditions and ff and mf conditions. When only one

variable's values change between conditions, the cause of any resulting change becomes more easily explainable.

In hypothesis 1 for example, the researchers focus on juror ratings of harasser according to gender of harasser while testing combinations of harasser and target gender. In hypothesis 1 it is hypothesized that the gender of harasser is a causal variable, and causal conclusions are drawn about gender of harasser in cross-gender cases (i.e. is the harasser guilty). In order to conclude anything about the construct 'gender of harasser,' gender of harasser would have to be the only variable differing between conditions. But gender of target also varies between conditions in hypothesis 1 because the two conditions being compared are a male harassing a female target and a female harassing a male target. Therefore any causal conclusion drawn solely about the gender of harasser creates confusion about whether the participants are basing the ratings on the harasser or on the target or on the combination of the two?

Researchers in this article were careful to refer to combinations when reporting the results but the conclusions created from these results could easily be misinterpreted when the specifying factor (i.e., that all results about harasser and target are in relation to the gender of the other person involved) is not clearly articulated.

Issue 2: To manipulate and communicate the gender of harasser and gender of target to participants, participants were exposed to vignettes describing the sexual harassment incident and photographs of the persons involved. Gender of harasser and target were, therefore, communicated via two avenues to the participants.

Photographs can be a complicated method of manipulation. The four photos used in this study, two of men and two of women, were tested to insure that the physical

attractiveness and age of persons in them did not significantly differ. All four people in the photos were White. However, what about other variables, such as facial expression, size of the person (photos included head and shoulders), style of dress, stereotypes about hair and eye color, and many other variables that might differ among the photos and thereby contribute to differences among the ratings in each condition. It is possible that photos may have contributed information, besides gender, to the participants' ratings, thereby complicating the attribution of differences among conditions to gender, or to the behaviors described in the vignettes.

Mono-method bias does not present concern here. Treatment was implemented via two methods: manipulation of harasser and target gender took place through the written vignettes and through photographs.

The researchers attempted to assess *reactivity to the experimental situation* by measuring the social desirability of participants. Their analysis suggested that social desirability presented no problems in the study.

Other threats to construct validity, including *mono-operation, treatment sensitive factorial structure, reactive self-report changes, experimenter expectancies, novelty and disruption effects, compensatory equalization, compensatory rivalry, resentful demoralization, and treatment diffusion* did not present threats to this study.

Applying the Regression Framework

Because the researchers conducted ANOVAs, they did conduct analysis within the general regression framework. But the above Elaboration Model analyses for this article make the difficulty in applying the regression framework apparent. Discussion of the lack of inclusion of 'gender combination' as a variable in the analyses and critiquing

the statistical analyses that were instead conducted takes me out of the bounds of the current goals of this dissertation. It is important to note, though, that it seems that the way in which data were statistically analyzed does not match the proposed hypotheses.

Summary of Results After Applying Each Model to Article 2

Applying all three methodological models to this study of sexual harassment resulted in differing perspectives on the validity of the causal conclusions. Rubin's Causal Model would deem no causal conclusion about gender as valid; the Elaboration Model did not fit well with the structure of the hypotheses or analyses in this study, thereby creating difficulty in using the logic of the EM to assess causal inferences; and the Validity Typology, while not identifying serious concerns with the internal validity of the conclusions (and therefore with the causal nature of these conclusions), did identify issues of concern to construct validity.

According to RCM, gender cannot be a causal variable, and therefore the causal conclusions about gender in this study are not valid inferences. Because all of the independent variables in this study were gender variables (gender of harasser, gender of target and gender of participant), this means that all causal inferences in this study would not be valid according to RCM.

This becomes somewhat complicated when we consider two seemingly contradictory issues: 1) gender qualified as both an 'S' and a 'G' variable in this study and 2) the necessary pieces to construct the causal effect equations were present for causal hypotheses about gender, which is an attribute. As chapter 13, the discussion chapter, will examine in detail, these contradictions can be explained by considering the

source of the attribute variable values (i.e., the unit supplying the outcome measures versus a stimulus separate from the units supplying outcome measures).

Surprisingly, the Elaboration Model proved quite difficult to apply to this study of sexual harassment. In the process of analyzing this study according to the EM's logic, the main problem arose when attempting to identify the original bivariate relationship and test variable in each hypothesis. This proved problematic because 'gender combination' of the harasser and target served as the main variable of interest, but did not exist as an actual variable for the statistical analyses. Subsequently, setting up the crosstabulation tables for testing a bivariate relationship partialled by a test variable could not be done within the EM's framework.

According to the Validity Typology, no problems existed for the internal validity of the study. Therefore, no problems existed for the causal nature of the conclusions. Unlike RCM, the VT finds no problem with using nonmanipulable variables as causal variables. However, the construct validity of this study suffered somewhat for two reasons. First, similar to the reason why the EM's logic did not suit this study well, the inclusion of two variables, gender of harasser and gender of target, both changing between conditions of comparison, creates confusion as far as what exactly contributed to participants' ratings of verdict, reward and perceptions. Was it the gender of the harasser? The gender of the target? The combination of the two? One cannot say because more than one of these variables differed at a time in the comparisons made.

In addition, participants were presented with photographs. Even though these photos were tested for differences in attractiveness and age and were all of White persons, so many other variables could differ among the photos, raising the issue of what

variables contributed to participants' ratings, and findings of differences among conditions.

Article 2 also highlighted some places where Rubin's Model in particular may benefit from added clarifications. Mainly, 1) the issue of attributes being excluded from the realm of causal analysis because a participant cannot be exposed to more than one level of an attribute became problematic, and consequently, 2) the match/mismatch between the causal variable equation and the causal variable definition requires further investigation.

CHAPTER 12: ANALYSIS OF ARTICLE 3:

ANDERSON, JOHN, KELTNER, AND KRING (2001), "WHO ATTAINS SOCIAL STATUS? EFFECTS OF PERSONALITY AND PHYSICAL ATTRACTIVENESS IN SOCIAL GROUPS"

The third article selected for this analysis was published in a 2001 volume of the American Psychological Association's Journal of Personality and Social Psychology. The article includes personality and attractiveness, which may or may not fit RCM's causal variable definition, and includes much causal language throughout. In this chapter, all quotes from Article 3 are accompanied by only a page number with the understanding that this is the article under analysis and the author and date are already known from the chapter title.

Three studies comprise this article. The first study provides a cross-sectional analysis of personality and social status in fraternity members. The second study replicates the first, but with a sample of sorority members. In both of these studies, participants filled out self-report measures about personality (the NEO), independent observers rated the physical attractiveness of participants on videotape and social status ratings were collected from peers as well as a counting of the number of positions each member held in his/her house.

The third study further investigates the personality-status relationship through a longitudinal study of both men and women. Participants in this third study were dormitory residents. Physical attractiveness ratings were collected the same way as in studies one and two. Personality was rated via self-reports as well as through objective ratings of videotape observers. Status measures were collected from peers and also from a

subset of participants who completed the same measures about themselves that their peers completed about one another. Data for the three studies were analyzed using correlations, regression analyses and ANOVAs.

Causal Language

All three studies concern the causal effects of personality and physical attractiveness on social status among college dormitory and fraternity/sorority students. Causal language abounds throughout this article (including in the title), but periodically intermingles with associative and predictive language, leading to the question of whether or not language was, at times, consciously chosen.

In the introduction, the researchers speak of the causal power of status, which is actually the outcome variable in their study. “Status attainment leads to a host of vital consequences for the individual. Research has shown that individuals’ status within their group influences personal well-being, social cognition, and emotional experience (citations)” (p. 116). Then they proceed to the causes of status: “...little empirical research has examined the personological origins of status” (p. 116).

At times in the introduction, personality is discussed both as associated with and as a cause of status: “A number of competing hypotheses may be formulated regarding the effects [*causal*] of Agreeableness. One possibility is that Agreeableness relates positively [*associative*] to status because...”(118). Causal language predominates, though, over associational and predictive language: “In the present studies, we focused on the Big Five personality dimensions (citations) as potential sources [*causal*] of status differences” (p. 118). “Finally, Extraversion itself is a valued characteristic and thus should lead to [*causal*] higher status in these social groups” (p. 118). The theoretical

premise of the article is stated as: “Status attainment is a function of [*causal*] both the individual’s personality and the group’s values and perceptions” (p. 118).

The research includes a mainly causal approach to physical attractiveness as well: “We also studied the effects [*causal*] of physical attractiveness on status in face-to-face groups.” (119) “We were also concerned that effects [*causal*] of personality on status attainment might be due to [*causal*] differences in attractiveness.”

This study of status attainment includes the variable ‘gender.’ The approach to gender includes a mix of causal and associational language to explain why males may not be comfortable reporting about their personalities: “Because [*causal*] high Neuroticism violates male gender norms governing the experience and expression of emotion, the relations between Neuroticism and status should depend on [*causal*] gender, with a stronger negative relation [*associational*] in men than in women” (119).

For the reporting of results and conclusions, at the end of each study (there were 3 studies) a ‘Results and Discussion section’ followed. All three ‘Results and Discussion’ sections contain mainly predictive language. The ending of the paper includes a lengthy ‘General Discussion’ in which both predictive and causal language are used to expand upon and explain the results of the study.

In the ‘General Discussion’ section, the type of language used to summarize the study varies. The beginning of the discussion seems to include more predictive language; as the discussion proceeds, the reader encounters more causal language, interspersed with predictive and associational language. The following two sentences appear in the first paragraph of the general discussion on page 126: “The primary aim of the current research was to examine whether the status ordering of face-to-face groups could be

predicted [*predictive*] from individual differences in personality traits and physical attractiveness.” “Moreover, our findings show which personality traits are important in shaping [*causal*] the status ordering of informal social groups.”

Applying Rubin’s Causal Model

This article presents challenges to Rubin’s Causal Model by clearly dealing in a causal way with variables that are difficult to classify according to RCM’s causal variable definition. How would RCM treat personality and physical attractiveness? What about the variable ‘time’ that enters in the longitudinal study?

S: Exposure Variable

Can participants be potentially exposed to different values of their own personalities? Although the measure of personality used in these studies is one that assesses personality dimensions as traits as opposed to states, i.e., enduring as opposed to changing personality dimensions, the researchers momentarily suggest the idea that personality is learned. This idea comes through in their discussion of why men and women may have differing personalities and why particular personality dimensions may matter more for men versus women (e.g., neurotic men are seen as less manly, while neurotic women are not judged as harshly).

In what context could personality be learned? If personality is learned, most of this learning probably takes place in early childhood. Is it possible to manipulate one’s exposure to personality? That depends on how much you believe that personality is nature versus nurture. If genetics determine personality, personality cannot be manipulated, and even to the extent that personality is not genetic, the stability of

personality traits may prove difficult to overcome, and possibly more so in some people than others (due to personality differences).

Can one be exposed to taking on different levels of physical attractiveness?

Physical attractiveness can change to a certain degree as a person wishes to change how s/he looks. A person can have braces, plastic surgery, or wear makeup to change appearance. Each of these modes for altering appearance leads to a different degree of change in physical attractiveness, some take years, some take just minutes. But how much alteration would be necessary to consider that a person's physical attractiveness actually changed from one level to another? How would other aspects of a person change in the process of changing attractiveness? Some aspects of physical appearance cannot be changed. The classification of physical attractiveness is not a straightforward one.

Personality and physical attractiveness may be variables for which a person could be exposed to different values, but these are not manipulable variables in the sense that a participant could be exposed to different values of the variables without a lot of ifs: if the participants were treated differently by their parents since birth, they might have different personalities; if participants could have grown taller as children, they might be considered more attractive. The list of 'ifs' is quite extensive. This line of thinking resembles the discussion presented earlier in this paper when Glymour suggests that if we were to go back to the formation of genetic material for a person with Down's Syndrome, we could potentially expose that person to not having Down's syndrome, a sequence of thinking involving a lot of 'ifs.' Practically speaking, personality and physical attractiveness do not fulfill the prerequisites for being causal variables in RCM. But their

classification will be questioned throughout this analysis. Classifying these variables as non-causal is not a cut and dry situation.

‘Gender’ (the researchers use the term gender; the measure used is one of biological sex) certainly does not qualify as causal. The introduction section of this paper, as well as the analysis for article 2 above, presented a thorough discussion of the reasons why ‘sex’ is not causal according to Rubin’s model.

How would RCM treat the variable ‘time?’ The model does not, in fact, touch upon the variable of time. In the third study of the personality-status article, participants’ personality and physical attractiveness were measured at time 1, and their status was measured at times 1, 2, and 3. Time is not an attribute of the participants, but it also is not a variable under the researcher’s control. Time naturally occurs. But it does not naturally occur in the same way that biological sex does. In fact, it seems impossible to think about participants *not* being exposed to all levels of the time variable. A measure may not be taken at one of the time points, but the participant is still exposed to that time point.

Given the unusual nature of the time variable, it is not considered to be a variable in the same sense that personality, attractiveness and status are variables. Many researchers, however, use time as a causal variable – longitudinal studies are often designed in which time is considered a main variable. Age, is sometimes used as the variable representing the passage of time. Age is not considered causal in Rubin’s Model. But the paper currently under analysis did not use age, it used time. Time’s lack of inclusion in Rubin’s Model renders it without classification at this point.

The preceding discussion concludes that personality, attractiveness and sex are not S variables in studies 1, 2, or 3. Do these variables classify as 'G' variables?

G: Subgrouping Variable

Holland and Rubin's use of the 'G' variable, a subgrouping variable for which the value could not have been different for units, does not quite fit the type of variables that personality and attractiveness are. Biological sex of participants does qualify as a 'G' variable. Personality and physical attractiveness cannot be categorized as causal without significant doubts. They receive this 'attribute' categorization with some reservations

'G' variables, in the way that Holland and Rubin describe them, are subgrouping variables, grouping participants into categories. Not all of the 'G' variables in this study are categorical, though. Personality was rated on continuous scales, as were ratings of physical attractiveness. Creating categories out of these continuous variables forms artificial distinctions between levels of personality and attractiveness.

To classify personality and attractiveness, one might consider the 'X' variable in Rubin's Model because 'X' is a continuous variable. However, 'X,' which represents a covariate in the model, can be a causal variable, which personality and attractiveness do not seem to be according to the tenets of RCM.

Perhaps to best classify personality and attractiveness, Holland's notation, which differs somewhat from Rubin's, and is not widely used in applications of the model, might be a better fit here. Holland (1986a) uses the notation 'A' to represent an attribute of a unit (participant); he does not use 'G,' but presents 'A' as the alternative when a variable is not an 'S' variable. When an 'A' variable is involved in a study, the conclusions drawn are associational, not causal.

Here, in this study about personality and status, personality and attractiveness have been classified as attribute variables. We cannot realistically imagine an individual being *exposed* to different levels of personality; changing one's personality requires more than a simple manipulation of exposure. We also cannot realistically imagine exposing an individual to a different level of attractiveness without likely changing other aspects of that individual (e.g., self-concept, self-esteem, etc.) in the process. These are continuous attributes, to be distinguished from categorical attributes, i.e., 'G' variables, and also to be distinguished from covariates, which may be both continuous and causal, i.e., 'X' variables.

Y: The Outcome Variable

'Y', the outcome variable, contains as many versions as there are levels of 'S'. In these studies, there are no 'S' variables meaning there cannot be more than one version of 'Y'. This is because 'Y' is the value of the outcome when the unit is exposed to a particular level of the independent variable. Because the independent variables are not ones for which a person could be exposed to more than one level, the model dictates that the value of the outcome, which is 'status,' could not have been any different than what it was for any given individual. The fact that only one version of 'Y' exists classifies results from studies 1, 2, and 3 as descriptive, not causal, results.

A number of hypotheses were posed in this paper. Because many of them were of the same type in studies 1 and 2, representative examples are selected for analysis below. The following hypothesis was posed: "...Extraversion should affect status attainment in face-to-face groups..." Using Rubin's symbols, a causal equation would not make sense because extraversion is not causal. The word "affect" in the hypothesis indicates the

researchers may have been considering extraversion as a causal variable, but in Rubin's Model extraversion qualifies as an attribute, classifying this hypothesis as descriptive. However, given that Rubin's notation does not account for a variable such as personality, a continuous and non-causal variable, the equations for the descriptive inferences cannot be demonstrated here.

As discussed above in the context of study 3, time does not fit into the classification of any variables, 'S,' 'G,' or 'X' in Rubin's Model. Therefore, to create the equations using Rubin's notation for the hypotheses testing time would be inappropriate. It is, though, important to note that in studies used as examples in the literature on Rubin's Model, time has not been dealt with as a causal variable. We cannot say that status at time 1 lead to status at time 2 because there is no condition of comparison for exposure to time 1. A causal effect results from the comparison of outcomes from multiple levels of a treatment variable.

X: Covariate

'Physical Attractiveness' was important both as an independent variable *and* as a covariate in these studies. For example, in Study 1 the correlation between physical attractiveness and status was significant, but when using attractiveness as a control variable (an X variable), none of the relationships between the personality dimensions and status changed. For example, after Study 1, the researchers concluded, "Overall, these findings suggest that physical attractiveness and personality are two truly independent pathways to social status" (p. 121).

Validity of Conclusions

To begin, any causal conclusions about personality, physical attractiveness and sex do not qualify as valid if we classify these variables as attributes. Sex clearly falls under Rubin's attribute category. But personality and attractiveness seem a bit too malleable to rule out of the causal realm completely. These variables may be malleable over long periods of time (and perhaps shorter periods for attractiveness), but we cannot imagine randomly assigning people to different personalities. As for the variable 'time,' its role in this study is somewhat unclear. It seems to play a supporting role, never in the foreground as causing change in status or contributing to the link between personality and status, but present in the interpretation of analyses.

Because the independent variables have been classified as attributes, there is only one version of the outcome variable 'status' in Studies 1 and 2 (study 3 is longitudinal). Only one value of 'Y' can exist for each participant, making the comparison with a counterfactual condition for an analysis of causal effects impossible. "Descriptive questions differ from causal questions in that there is no implied comparison of the values of an outcome variable under different levels of an experimental manipulation" (Holland & Rubin, 1983, p. 19). Personality values do *not* indicate levels of a treatment. Attractiveness seems more manipulable than personality in the sense that smaller events lead to change in attractiveness. Certainly with today's technology, experimental studies could be created to test a wide range of changes in physical attractiveness, but the experimenter would need a precise definition of what constitutes an actual change in one's level of physical attractiveness, and likely that change would entail simultaneous changes in other aspects of the participants.

Given that the variable ‘sex’ played an important role in this research article, was it acknowledged that sex was acting as a proxy variable, or was sex used in a causal sense indicating that differences found were a result of sex. Momentarily it seemed that the idea of proxy was touched upon, for example: “In combination, these findings offer considerable support for the hypothesis that gender norms about negative emotion are involved: ‘Real’ men are not supposed to feel and act afraid, sad, guilty, or vulnerable, and men who violate these gender expectations are less likely to be granted high status in face-to-face groups” (p. 127).

Was sex viewed as a proxy for societal gender norms and gender expectations? Meaning did the idea enter the analysis that sex represented socially constructed norms and expectations rather than simply sex differences? Although I like to see this as a proxy, I think the researchers use gender norms and expectations as *explanations* for gender differences. Proxy variables differ from explanations. A proxy acknowledges that unmeasured variables acted through the measurement of the proxy, whereas an explanation occurs as a reason why a relationship exists between variables without acknowledgement that a measured variable *represents* other related variables.

Applying the Elaboration Model

Across the three studies in this research, there were four hypothesized causes: personality, physical attractiveness, sex, and time. The effect of interest was social status. In the first two studies, sex was controlled for via design; the first study included only men and the second study included only women. In these two studies, the main cause was personality and physical attractiveness was of secondary interest. In the third study sex and time enter as additional variables.

In Studies 1 and 2, the main bivariate relationship existed between personality and status. Physical attractiveness entered the analysis in three ways: The researchers formed hypotheses about the relationship between physical attractiveness as causing a certain status level, as well as physical attractiveness being controlled for in the relationship between personality and status, and also the idea that physical attractiveness might interact with personality to affect status.

After first testing the relationships between each of the Big Five personality dimensions and status through correlation analyses, physical attractiveness was controlled for via partial correlations. In separate analyses, the interaction between physical attractiveness and each personality dimension was entered into multiple regression analyses. No significant effect occurred in either study for the interactions of attractiveness and personality.

Although personality was the main independent variable of interest in the study, hypotheses questioned the possibility of interactions between multiple personality dimensions. For example, the interaction of agreeableness and extraversion on personality was tested. The interaction's placement in a section about the role of agreeableness in determining status suggests that extraversion acted as the test variable in this analysis, while agreeableness was the main independent variable of interest.

Study three was more complicated than these first two studies. In study three, personality was still the main cause of interest, physical attractiveness was included and status remained the effect. But two additional variables entered the analysis: sex and time. This third study included both men and women measured at three points in time to test the stability of status.

Data for men and women were analyzed separately at two different time points, four and nine months after the first measurement two weeks into the school year. Both sex and time qualify as test variables, making three test variables in this third study because physical attractiveness remained a covariate.

Study three also included some analyses aimed at eliminating possible alternative explanations separate from the main analyses. These additional analyses exemplify the heart of the Elaboration Model's mission – to identify and test variables that will aid in understanding a relationship. For example, a 2 x 3 ANOVA was run to make sure that “there was no consistent sex-of-target difference across time” (p.124) with sex as a between factor and time as a within factor, and status as the outcome. Here, sex was used as a test variable and the desired resulting partial relationships fall under ‘replication.’

Next, “We compared the attractiveness ratings from the fraternity and sorority samples with the dormitory men and women in a 2 x 2 ANOVA, with sex and living group as between-subjects factors.” Here the outcome was attractiveness, and the samples used in the study were analyzed by sex to eliminate sex as an explanation for attractiveness ratings. They found a main effect of living group (Greek vs dorm), but not of sex or an interaction between sex and living group.

Even though the researchers only found a significant correlation for extraversion with status, they tested the interaction of physical attractiveness with each of the Big Five as they relate to status. Therefore, they continued to investigate the relationship regardless of the fact that no significant relationship resulted from the bivariate test (extraversion and status). Such a decision can produce meaningful results, uncovering a possible suppressor variable or distorter variable.

When considering the time order of test variables, it appears that personality and attractiveness were hypothesized as antecedent to status, “Can we show that preexisting personality traits and physical attractiveness predict status in newly formed groups?” (p. 123). But is attractiveness antecedent to personality or intervening between personality and status? Someone who has a very outgoing personality may dress differently, and fix himself or herself up in a different way than a person who is more introverted or neurotic – and the way one chooses to present oneself certainly affects how others perceive one’s physical attractiveness level. At the same time, someone who looks attractive may feel more comfortable with extraverted behavior and feel less anxious about social interactions, making attractiveness antecedent to personality. So, I am not sure where this variable falls in the time order. Perhaps it is neither antecedent nor intervening, as the EM accepts variables that do not fall into either domain.

The researchers tested specifically for *specification* relationships (by testing interactions in studies 1 and 2) but they found nothing here. In Study 1, the type of relationship resulting was *replication*, because after entering physical attractiveness as a control variable, nothing changed in the relationship between personality and status.

Interestingly, in Study 2, there was no relationship between the test variable (attractiveness) and the outcome variable. But they still controlled for it. In Study 2, the resulting relationships were also *replication*.

This study fits well within the ideas of the Elaboration Model. The researchers began with bivariate relationships, then added variables to investigate how the original relationship would change after new variables entered the analyses. They began by testing the sexes separately, and then conducted an analysis with sex as a test variable

(both sexes were included in the analysis). As well, they established the relationship between the original variables at a single time point, and then expanded this to investigate how personality and physical attractiveness predict status at 3 time points.

Applying the Validity Typology

By reading this research article, it appears that the researchers followed the logic of the Validity Typology by identifying and testing plausible rival hypotheses. In the design stage, plausible threats to internal and construct validity lead to deeper consideration of operationalizations, multiple methods, and the need for multiple studies.

Studies 1 and 2 were cross-sectional designs with no actual treatment and study 3 entailed a longitudinal design with measurements of outcomes taking place when participants reported on their personality, and observers rated the participants' physical attractiveness levels. Studies 1 and 2 assessed personality, attractiveness and status at the same time, while study 3 measured status at three times. These studies included no treatment, no manipulable variables. The Validity Typology does not provide notation to match these studies. The notation of 'X' and 'O' applies to a study in which a treatment is applied and then an outcome is measured. 'X' does not represent the independent variable here because in this case the independent variable occurred through participants' self-reports of personality, which is naturally occurring (and of course, includes the nature vs. nurture debate).

The groups used in studies 1, 2, and 3 existed prior to the research; no form of assigning participants to conditions took place in those studies. It seemed as though the researchers considered threats to validity because they took methodological measures to eliminate particular threats, as discussed below.

Internal Validity

To rule out the threat of *ambiguous temporal precedence*, that is, confusion about the temporal order of variables, the researchers carried out a longitudinal study. The first two studies assessed personality and status simultaneously making it very difficult to decipher which variable was the cause. In Study 3, assessing personality at time 1, and status at times 1, 2 and 3, provided some evidence to support personality preceding status attainment. But, given that these variables were not under the researcher's control, we do not know if the temporal order may be due to a variable antecedent to personality and status.

Maturation possibly contributed to lower internal validity in Study 3. After all, the participants consisted of new college students, and outcome measures collected over the first year of college could definitely include the influence of psychological changes taking place during such a transitional life stage.

Instrumentation may have been an issue over time as status measures were repeatedly collected from peers in Study 3. But given that status measurements from all peers were averaged for each participant, if some peers but not others changed their measurement system over time, this may have mitigated the threat to the reliability of measurements over time.

Additive and interactive effects of threats to internal validity may be relevant for an interaction between *maturation* and *instrumentation*. As participants got to know more and more students, and had more interactions with students over the year, likes and dislikes may have had time to form and affect the way in which some students rated and were rated. In addition, participants' ability to judge others' personalities may have

improved over the year after having met so many new people in such a short period of time.

Selection, history, regression, attrition and *testing* did not present plausible threats to the studies.

Construct Validity

Regarding construct validity, researchers of this study were careful to match the operationalization of status with the ‘status’ concept as they defined it, ruling out *inadequate explication of constructs* for the outcome. A thorough review of relevant literature highlighted three components of the variable ‘status,’ and only instances of ‘face-to-face’ status were of concern for the study.

To ensure that the operationalization of the outcome variables did not capture more than one construct (i.e., *construct confounding*), defining ‘status’ involved distinguishing it from ‘leadership’ and ‘popularity,’ and identifying respect, influence and prominence as key components of status.

To prevent settings from being over or underrepresented (settings being an aspect Shadish et al. bring more attention to in the 2002 discussion of construct validity), the researchers specify the construct under study involves only face-to-face status.

In the case where *mono-operational bias* was recognized as a plausible threat, which was male self-ratings of neuroticism, the researchers included a second measurement – that of an observer’s rating of the participant’s neuroticism. This second measure provided a more objective measure of male neuroticism, to prevent bias in the measure from males not wanting to self-report neuroticism due to violating social norms.

Researchers also used two personality inventories, the NEO and BFI. Two inventories were used because the NEO results in a greater correlation between agreeableness and extraversion than the BFI.

Mono-method bias was not problematic. Participants reported their own ratings on personality measures; for Neuroticism, independent observers also contributed ratings in Study 3. Independent observers coded physical attractiveness. Status was reported by peers, and assessed through life outcome data regarding the number of positions each participant had held. Therefore, several data collection methods were used. These multiple methods insure that results embody the relationship between constructs, not methods.

Other construct validity threats not plausible for these studies include *confounding constructs with levels of constructs, treatment sensitive factorial structure, reactive self-report changes, reactivity to the experimental situation, experimenter expectancies, novelty and disruption effects, compensatory equalization, compensatory rivalry, resentful demoralization, and treatment diffusion.*

Applying the Regression Framework

All variables included in the analysis of Studies 1 and 2 were continuous variables. Relationships between the five personality dimensions and status, and physical attractiveness and status, were tested through correlations. Additional hypotheses of controlling for attractiveness, testing curvilinear relationships, and interactions were tested through multiple regression.

The researchers controlled for physical attractiveness by entering it into a partial correlation with each personality trait that was significantly related to status. They tested

for interactions between each of the five personality traits and physical attractiveness by entering an interaction term of each personality dimension with attractiveness into a multiple regression equation such as:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2$$

where X_1 is a personality trait, X_2 is attractiveness, and $X_1 X_2$ is the interaction of trait & attractiveness.

On the basis of competing ideas that “Both too little and too much Agreeableness will hurt individuals’ chances for status attainment, whereas moderate levels of Agreeableness are best for status attainment” (p.118), a quadratic relationship was tested by entering a squared term into the regression equation. Squaring a term tests for a curvilinear relationship between a predictor and an outcome, as opposed to a non-squared term which tests for a straight-line (linear) relationship between a predictor and an outcome. Squaring a term represents the interaction of a variable with itself.

$$Y = B_0 + B_1(\text{agreeableness}) + B_2(\text{agreeableness} \times \text{agreeableness})$$

In Study 3, correlations tested the stability in social status between time 1 and time 2, and between time 2 and time 3, separately for men and women, and for both sexes combined. Then, each of the Big Five was correlated with status at time 2, and with status at time 3, separately, and separately by sex to assess the answers to such questions as “Did

Extraversion, measured at the first assessment, predict status 4 and 9 months later?" (p. 125). Partial correlations controlling for physical attractiveness followed.

One issue guiding Study 3 was temporal stability in status: "Measuring status at three different times also allowed us to explore an important issue in the development of status, namely its stability and change over time" (p. 123). Separate correlations for status between time 1 and time 2, and between time 2 and time 3 indicated the stability during two points in time. For the men, stability from time 1 to 2 resulted in a correlation of .80, and from time 2 to time 3, in a correlation of .89. For women, the correlation from time 1 to 2 equaled .41, and from time 2 to 3, .88. All four of these correlations were significant.

From the analyses, the researchers concluded: "In combination, these findings demonstrate that the hierarchy in men developed and stabilized rapidly, whereas the hierarchy in women stabilized more slowly" (p. 125). No test was conducted to examine whether or not these two correlations for the men were significantly different though. Was there significant change in status between times 1 and 2, and times 2 and 3 for men? Using the Fisher Z transformation to compare two correlation coefficients would provide an answer to the inquiry of whether or not stability or change in status existed.

Summary of Results After Applying Each Model to Article 3

Interestingly, the application of the three models offered drastically different considerations about the causal conclusions from this article, and contributed some important areas of potential change in RCM. Applying RCM led to multiple areas of confusion and difficulty. Applying the EM and VT proved to be quite straightforward with this article.

The variable 'time' presented issues of its own. The nature of time does not fit well with the idea of potential exposability, and RCM offers no appropriate classification for 'time.'

RCM's causal variable definition created complications for the 'personality,' 'physical attractiveness,' and 'time' variables. Personality and attractiveness did not fit with the requirements of the causal variable definition. Yet they did not fit perfectly with the attribute definition, either.

In addition to the difficulty with the causal variable definition, problems applying RCM's notation also arose because these variables were continuous, and RCM does not provide notation for non-causal continuous attributes. Notation presented only by Holland (1986a) best fit the nature of these two variables.

The logic followed by the researchers fit well with both the EM and the VT. The causal conclusions about personality and attractiveness that presented problems for RCM presented no problems for these other two models. Some awkwardness was encountered, however, when attempting to apply the VT notation to the studies in this article. The continuous and nonmanipulable nature of personality and attractiveness did not fit well with the experimental notation of the model.

CHAPTER 13: DISCUSSION

The purpose of this dissertation was to examine how the application of Rubin's Causal Model (RCM) contributes to the practice of inferring causal conclusions in social science. Interestingly, applications of this model to published research created a symbiotic relationship: Research from education, organizational psychology and social/personality psychology provided examples challenging extant ideas about RCM by locating places where the model could be strengthened, and RCM identified areas for needed clarifications in the methods we currently use.

Analysis of each of the three selected articles presented different challenges to Rubin's Model. All three articles included variables that RCM classifies as attribute variables, thereby creating problems for causal inferences according to Rubin's Model.

Analyses resulting from the use of the causal equation and causal variable definition created discrepancies with conclusions arrived at from the perspective of the researchers in the three articles.

Although not clashing with the conclusions drawn by the researchers, the tenets of the Elaboration Model (EM) at times did not mesh with the way in which the logic of an article, such as the articles about parenting and school readiness, and sexual harassment, proceeded. Applications of the Validity Typology (VT) presented problems for the parenting and school readiness article, and sexual harassment article conclusions. The third article, about personality and social status, on the other hand, seemed to use the VT as a guide; the authors carefully considered threats to each of the four types of validity and, through design, eliminated many of these threats.

A main goal of this dissertation was to create a list of issues for researchers to consider when designing and interpreting the conclusions from a causal study. In pursuit of that goal, this discussion consists of three main sections.

First, a brief summary of the analysis of causal language concentrates on the difficulty in identifying clearly causal language and the need to carefully consider what a conclusion communicates, both directly and indirectly.

Second, applications of RCM, the EM and the VT resulted in a number of comparison issues for discussion. Comparison of the three models' approaches to various methodological issues creates the opportunity to discover places where RCM both aids and hinders the process of causal inference. Comparisons were made among the models in earlier chapters; here the focus will be on the unique perspective of the EM and VT, perspectives that RCM did not bring to the articles. Table 12 summarizes comparisons among the three models.

Third, after summarizing the contributions of EM and VT to the analysis of each article, the focus will turn to RCM and areas of needed clarification, for the model itself and for social scientists to consider. In-depth discussions about the aspects of RCM that presented problems in the analysis create the final pieces needed to assemble a list of guidelines for researchers to consider when designing and interpreting causal research. These guidelines provide knowledge not only to the methodological practices in social science, but also to the content (e.g., attributes) that researchers choose to study.

Language

Examining the causal language in research articles provided indications of the way in which researchers hypothesized and interpreted relationships among variables in a study, although, one wonders how much purposefulness went into choosing *how* to word ideas in these research articles. At times the same idea was expressed through both associational and causal language. A relationship that is causal can also be described as associational, but an associational relationship is not necessarily causal. A causal relationship may be described as a predictive relationship, but a predictive relationship is not necessarily a causal one.

The acknowledgement that researchers may not purposefully choose language in published research conjures up concern at many levels. Most importantly, the language of hypotheses or conclusions carries implications for how another researcher or policy maker uses the study results, or cites the work to shape other research. Policy people are interested in manipulable variables – those that can be changed. And they are interested in for whom, for which subgroups of the population, manipulation of certain variables will make a difference (e.g., moderation). It is not difficult to imagine situations in which a person needs to know the results of a study and therefore trusts the conclusions presented in an article without detailed consideration of the method or variables used to arrive at those conclusions. Such situations highlight the imperative need for researchers to insure the validity of their own causal conclusions.

Given that we are taught to report regression results using predictive and associative language, perhaps there was a concerted effort not to interpret regression

results in a causal way in the results section, but such careful attention was not paid in the discussion. The hourglass shape approach to writing an empirical research article (Bem, 1987) recommends that the end of a discussion section should report ideas stemming from the generalization of findings, ideas that persons other than psychologists would be interested in reading. Perhaps the desire to emphasize the usefulness and applicability of one's research leads to sweeping conclusions not fully supported by empirical evidence.

The lack of purposeful causal statements contributed to the difficulty in identifying clear causal conclusions. That is, at times statements would imply causal relationships without consciously making the causal connection, statistically and conceptually. This phenomenon occurred in the relationship between time and status, and personality and time and status in the final article analyzed. It also complicated the interpretation of conclusions in the sexual harassment article, and conclusions about ethnicity in the parenting and school readiness article.

Elaboration Model

Unlike RCM, the EM provides no definition for a causal variable, so analysis of conclusions with the EM centered on the match between hypothesized relationships, order of variable importance, the use of test variables, and the resulting partial relationships. The Elaboration Model led to uncovering inconsistencies between relationships hypothesized and relationships actually tested in statistical analysis.

Particularly in Article 2, difficulty existed in interpreting the statistical analyses conducted in terms of the relationships put forth in hypotheses because the analyses,

unlike the hypotheses, did not include ‘gender combination’ as a variable. Article 1 also contained hypothesized relationships translated in a somewhat mismatched manner into analyses, with ethnicity and income hypothesized as both main independent variables and as test variables, but analyzed mainly as test variables, and tested moderation incorrectly.

Because the EM says nothing about what is and is not a causal variable or how to calculate a causal effect, ethnicity, sex, personality and attractiveness posed no threat to causal conclusions under applications of the EM. Also, there was no need to distinguish among different types of variables, ‘S’ or ‘G’. This lack of distinction diverts attention away from considering the suitability of a variable for causal interpretations but also allows the Elaboration Model to offer its unique contribution of classifying types of resultant relationships, which is important for furthering the development of theory. The causal part of the EM model is in asking what other variables may be responsible for clarifying why two variables are related, again contributing to informing theory.

The EM, unlike RCM and the VT, recognizes the importance of nonsignificant as well as significant relationships. Perhaps because lack of statistical significance leads to un-publishable research we often overlook nonsignificant results. But when a researcher finds no significance in a bivariate relationship, the EM suggests adding theoretically relevant variables to pursue a possible suppressor or distorter relationship. The inclusion of variables such as the ones in the articles comprising this analysis: ethnicity, gender, and physical attractiveness, might allow a

researcher to realize that a relationship does exist, albeit a more complicated one than originally hypothesized.

As well, the partial relationship ‘replication’ indicates the importance of considering when relationships do not change in the presence of a third variable. Research benefits from knowing which variables do, and do not, matter in relationships. The partial relationship classification recognizes the contribution of discovering the lack of a relationship to furthering knowledge.

In a structural sense the EM guided an analysis of logic for each article studied here, meaning it required consideration of the order of thinking about variables, the order of variable importance in an analysis and the decision about how to label results to further theoretical understanding (e.g., partial relationship classifications). In Article 1, for example, application of the EM entailed understanding the order in which variables were tested in analyses according to the relationships hypothesized to exist. Using the EM led to the discovery that variables were not entered into the analysis in the same order as they were hypothesized to relate. This led to questioning whether or not moderation was correctly tested, and if the roles of ethnicity and income were correctly assessed in relation to parenting and school readiness.

In Article 2, the EM brought attention to the fact that the hypotheses focused on variables (cross-gender and same-gender) that were never even variables in the analyses. The cross-same distinction occupied the focus of every hypothesis, but was never entered as a single entity into the analyses. This made it difficult to not only understand the results, but to interpret them accurately.

In addition, the EM's attention to labeling test variables as antecedent or intervening led to the realization, with physical attractiveness in Article 3, for example, that some variables are not well understood, and therefore present problems for drawing accurate conclusions. In Article 3, physical attractiveness could have been antecedent as well as intervening. But in the EM, these two classifications lead to possibly differing relationships (e.g., explanation vs. interpretation). To prevent this problem, researchers could more explicitly explain the role of a test, or control, variable in an analysis, and its relation, statistically and conceptually, to the other variables in an analysis.

Use and interpretation of test variables within the EM leads to serious consideration of the mechanisms at work in a causal relationship. By inquiring beyond the bivariate relationship, the EM encourages researchers to give thought to other variables that may help understand a relationship. Whether the variable proves unimportant for the relationship, or helps to specify under what conditions the relationship exists, or even makes the original relationship disappear (i.e., explanation), the inclusion of test variables creates the basis for furthering theory.

For the reasons stated above, researchers would benefit from following the logic of the Elaboration Model, from the inception of a research idea through the formulation of hypotheses formation, all the way through to the process of drawing conclusions.

Areas for Future Work

One area of possible clarification in the EM model itself refers to the statistical analyses appropriate to test for each type of partial relationship. The model

was founded using a series of contingency tables. The use of these tables demonstrates relationships between the original two variables, and then the original variables partialled according to a third variable.

What is the best way to statistically test for explanation, interpretation, replication, specification, suppression or distortion? What are the best tests for these relationships when the variables are categorical versus continuous? How does a researcher judge whether or not the partial relationships are weaker than or stronger than the original? The model does not offer such details for a researcher who wishes to implement it.

For example, in Article 1, the researchers included ethnicity and income as moderating variables, thereby testing the roles of these test variables through interactions. When the interactions did not result in statistical significance, would this be interpreted as replication? I phrase this as a question because controlling for these variables would also create a route to test replication, and the results of this route may differ from the results of the moderation approach. Specifying the statistical analyses appropriate to test the EM's partial relationships would add clarification to applications of the model.

Validity Typology

The VT and EM test rival hypotheses, albeit in different ways. EM focuses on measurable variables that might explain the relationship between variables, while VT focuses on design and measurement issues that may explain that relationship. The VT explicitly searches for alternative explanations, while the EM provides tools for researchers to incorporate additional variables when possible. Cause is a main

concern for the VT, while explanation is the main concern for the EM. RCM does not place much concern on the issue of rival hypotheses, although its assumptions (e.g., independence, SUTVA) in a way enter the rival hypothesis realm.

The VT and RCM directed attention at the temporal order of the variables, albeit in different ways. The VT considers the temporal sequence of the independent variable in relation to the dependent variable and questions whether or not the order could occur in a different direction. Which variable is the cause and which is the effect? RCM considers temporal order when classifying a covariate; a variable that has been affected by the treatment variable cannot be used as a covariate.

The VT also highlighted several aspects of study design untouched by RCM and the EM. The EM considers alternative variables that could have been tested to explain a relationship, but unlike the VT it did not lead to thinking about structural elements of the research. RCM rarely touched upon any structural elements, other than assignment mechanism for participants. For RCM, assignment mechanism captures the heart of causal thinking.

RCM and EM direct attention to the hypotheses to articulate the relationships under study. In Articles 1 and 2 the Validity Typology focused my attention on how variables were conceptualized and then operationalized according to past literature and current hypotheses. Neither the issues of variable operationalization and construct confounding with income and ethnicity in Article 1, nor the confounding of multiple measures of gender or the use of photographs differing along multiple dimensions, were ever broached by RCM or EM.

Applying the VT required returning to the ‘participants’ and ‘procedures’ sections of these articles to pay attention to how researchers structured the study. The VT’s structure forces one to consider the exact study design, whether it be cross-sectional as in studies 1 and 2 of Article 3, or longitudinal as in study 3. Certain threats to validity become readily apparent after articulating the study design. Of note is the fact that the VT model may be easier to implement at the design stage than RCM because RCM is not as developed, it is a newer model. RCM is much more general and abstract than the VT.

Unlike RCM, the VT did not problematize the issue of causal variables needing to be manipulable but its notation of ‘X’ and ‘O’ proved difficult to apply to the independent variables in Articles 1 and 3. The VT turned attention to considering the lack of treatment in the context of defining what ‘X’ represented when applying the VT’s notation system to these articles.

An important difference between VT and RCM is the consideration of construct validity. Construct validity occupies a large part in the Validity Typology because confounding constructs present critical problems to the inference of causal conclusions. No aspect of RCM focuses on construct validity. It is only raised when Holland (1986a) and Rubin (1986) explain why attributes cannot be causal, (i.e., we would not be able to change an attribute because in doing so we would also change other things about a person). RCM, as discussed earlier in this dissertation, emphasizes internal validity, seemingly at the expense of construct validity.

The lack of construct validity concern in RCM comes across clearly in Article 2. The VT offered a way of thinking about the hypotheses and conclusions that RCM

did not. Because those hypotheses involved cross-gender conditions, ‘gender of the harasser’ was not the only variable differing between conditions, ‘gender of target’ also differed between conditions. An assumption identified here by the VT was that harasser was being rated independently of ‘target gender,’ or was more important than ‘target gender’ in the minds of participants.

In RCM terms, the equation necessary if gender of harasser was the cause of interest would be:

$$(Y_{mm} + Y_{mr}) - (Y_{fm} - Y_{fr})$$

Granted, these researchers’ hypotheses only dealt with cross-gender situations, but because of that, we cannot conclude anything clear about harasser gender, only about the combination of harasser gender and target gender. Any conclusion about harasser gender would be confounded by target gender, as the VT points out.

Areas for Future Work

Even though the Validity Typology specifically covers nonrandomized research designs, and designs lacking control groups, the model does not offer as much guidance for survey research (i.e., correlational research), naturally occurring ‘treatment’ variables (i.e., manipulable vs. nonmanipulable independent variables) or continuous independent variables. Difficulty applying the model’s ‘X’ and ‘O’ notation highlights these areas of possible clarification for the model. The comprehensiveness of the VT, though, makes one wonder whether or not these issues are covered by the model and the generalizations and applicability to the issues mentioned above just do not come across clearly.

Rubin's Causal Model

Rubin's causal model offers its own unique perspective on the analysis of causal inferences. Rubin does not take credit for the underlying principles of his model, which he comments have been around in many aspects of research practice for many years (Little & Rubin, 2000). But from those principles he has built an interesting network of ideas to approach the topic of cause in research. Most existing applications of his model do not concern social science disciplines, but concern disciplines such as statistics, economics, and clinical trials in which treatments are the focus of study.

In Holland & Rubin (1983), studies are discussed in which treatments do not exist (e.g., Lord's paradox) and situations are analyzed in which variables are not classified as causal. Attempts have even been made to understand, in the context of RCM, the study of attribute variables about which researchers desire causal answers (e.g., Kaufman & Cooper, 1999). Rubin explicitly communicates that the model applies to nonrandomized as well as randomized studies. But in some disciplines, such as social sciences, nonrandomized studies mainly consist of groups created on the basis of attribute (non-causal) variables. Rubin would classify these as descriptive, not causal studies. But descriptive conclusions may not carry as much weight as causal conclusions, especially for policy formation.

Problems arise when applying this model in the social sciences. So many of the variables that social scientists study cannot be classified as causal variables in RCM. Yet studies using these variables do include causal conclusions about these variables. What is the danger in inaccurately inferring a causal relationship? At the

least, it starts a chain of misunderstanding the relationships among variables. At the most, it leads to political agendas and policies inaccurately targeting ‘causes’ of societal problems, denying people the opportunities they may need for equitable treatment in society.

Below I have highlighted areas in need of further consideration that arose during my application of RCM to the three articles under study. These issues fall into two main categories: 1) causal effect equation, and 2) causal variable definition.

Causal Effect Equation

Applying the causal effect equation requires first labeling variables as ‘S,’ ‘G,’ or ‘X.’ The above analyses demonstrate that labeling variables accordingly can be quite complicated. Interpreting the resulting equations also begets complications with continuous variables, interactions and abiding by the causal variable definition. The issues arising around the causal effect equation involve: 1) distinguishing between ‘S’ and ‘G’ variables and acknowledging the need for additional notation; 3) extending the extant frame for interpretation of the equation including questioning the mismatch between classifying causal variables and classifying control variables and exploring the consistencies and inconsistencies between the causal effect equation and the causal variable definition.

S Versus G Variable

Rubin’s Model distinguishes between ‘S’ and ‘G’ variables. ‘S’ denotes a variable to whose conditions the participants are exposed. ‘G’ signifies a subgroup to which a participant belongs. The value of ‘S’ for any participant could be potentially different than what it is; ‘S’ is manipulable. The value of ‘G’ for any participant could

not be different than what it is; 'G' is not manipulable. 'S' represents the levels of a causal variable; 'G' represents levels of a non-causal variable.

The delineation between these two types of variables seems quite straightforward and the differences between them may seem easily distinguishable. But as demonstrated by the many discussions above about classifying a variable as causal versus non-causal, serious thought must be given to deciding between the 'S' and 'G' classifications. A variable's classification as an 'S' variable opens the door to causal inferences, whereas a variable's classification as 'G' 'limits' one to descriptive conclusions.

A clear example of this arises in Article 2. The variable 'gender' is not a uni-dimensional one in this study. When 'gender' is an attribute of the study participant, it is classified as a 'G' variable; when 'gender' is an attribute of the person in the vignette, it becomes an 'S' variable; when 'gender' is an attribute of the person in a picture shown to participants, it is an 'S' variable. The difference, as will be discussed in the causal variable section below, concerns whether or not the attribute belongs to the unit under study, or to an object or scenario presented to the unit under study.

In Article 1, the researchers hypothesize 'ethnicity' as what could be considered an 'S' variable in one hypothesis, but as a 'G' variable in another hypothesis. The way in which the researcher hypothesized the role of ethnicity mattered for whether or not the variable was classified as 'G' or 'S' according to the researcher (not according to RCM because RCM does not classify ethnicity as causal).

According to RCM the distinction between ‘S’ and ‘G’ variables matters because ‘S’ variables are causal, whereas ‘G’ variables are not. Many variables in the social sciences might at first glance be categorized as ‘G’ variables: variables defining preexisting groups (groups not formed by the researcher such as from demographics and personality). These variables may not be manipulable at face value, either because it is impossible or because it would never be ethical.

Additional notation needed.

As demonstrated in Article 3, the S vs G notation does not cover all independent variables (that are not covariates). Personality and physical attractiveness were continuous variables, not classified as ‘S’ variables, but also not fitting into the definition of a ‘G’ variable. Holland and Rubin (1983) never used the ‘G’ notation with a continuous attribute variable. As ‘G’ is defined as a subgrouping variable, perhaps the concept of subgrouping could be expanded to include each level of a continuous variable, thereby allowing G to take on any value between 0 and 100, for example. In Article 3, though, without notation explicitly matching the structure of a continuous attribute, Holland’s (1986a) general attribute notation, ‘A’ was implemented.

In addition, the variable ‘time’ in article 3 could not be accounted for by the model’s notation. The issue of ‘time’ may run deeper than simply a notational problem, though. As explained in the analysis of Article 3, time is sometimes used as a causal variable, but thinking about time according to the potential exposability idea creates some confusion. No researcher can control the passage of time, but the researcher can control at what time a participant may be exposed to a treatment.

Time's levels are not mutually exclusive, either. It is not possible to consider exposing someone or not exposing someone to time. Its role in RCM should be further explored.

Interpreting the Equation

The original definition of a causal effect, $Y_t - Y_c$, becomes more complicated when working with multiple variables or continuous variables. In such situations, how does one go about applying this equation in an interpretable and useful manner?

This issue arose in the analysis of parenting style in Article 1, as well as for personality and attractiveness ratings in Article 3. In Article 1, it became difficult to decide what would count as an appropriate counterfactual condition for parenting given that parenting variables were rated on Likert scales. It was decided that the *functional form* of the relationship between the continuous independent variable and the dependent variable (e.g., the shape of the regression line) must be known in order to make meaningful causal comparisons.

In the literature, discussions of Rubin's Model do not include examples of interactions between 'S' and 'G' variables. Holland and Rubin (1983) provide the example of calculating causal effects in subpopulations, "When subpopulations have been defined using G , it is natural to want to estimate a causal effect in each subpopulation" (p.20). By definition an interaction communicates the idea that the effect of one variable (treatment versus control) *depends on* the level of another variable (possibly a 'G' variable).

Practically, it makes sense to consider examining the interaction of a 'treatment' (S) with a subgrouping variable (G) and this is common practice in social

science. When understanding the implication of an S*G interaction, one would say that 'S' causally relates to the outcome, 'G' is associated with the outcome, and the interaction of the two variables means the effect of 'S' on the outcome depends on the level of 'G.'

But with the concept of moderation, for example, in Article 1, the interaction constituting this moderation consisted of an 'S' variable (parenting) and a 'G' variable (ethnicity). Moderation in that discussion [and in Baron & Kenny's (1986) widely accepted explanations] was considered a causal concept. This raised concerns about the use of 'ethnicity' according to the RCM framework. What, then, becomes the correct interpretation of moderation? Is it causal, regardless of the variable types included (S vs G)? Or is it only interpreted as causal when no 'G' variables appear in the interaction?

Control variables.

One final issue in interpreting the causal effect equation is the use of a covariates, or control variables. When the covariate, as in Article 3's use of physical attractiveness, is not a causal variable, how is its use interpreted in the equation? 'G' variables are a type of covariate (Holland & Rubin, 1983). Covariates and control variables *are* causal concepts, as discussed in chapter's 2 and 6 of this dissertation. Physical attractiveness may not be the best example here because its classification as non-causal is murky. However, recall the many cases in which race and sex are used as control variables. How, according to RCM, would the role of these attribute control variables be understood?

Match between causal equation and causal variable definition.

In Article 2 about sexual harassment, it became apparent that the causal variable definition and equation are not as redundant as they seem. The causal variable equation in this article could be completed by the comparison of any two conditions in that study, but all conditions were created through the manipulation of gender variables. Most of the equations written above, in the analysis for Article 2, meet the conditions necessary for a causal effect analysis – they include manipulated variables and participants could be exposed to all conditions included in the variables. But when the variables are judged according to RCM's causal variable definition, these equations become nonsensical because they involve attribute variables.

Causal Variable

A good portion of this paper has been dedicated to examining and questioning the definition of a causal variable offered by Holland and Rubin's research on RCM because without a causal variable, a discussion of causal effects is moot. As outlined in the introduction, the classification of causal versus attribute variables makes for sticky distinctions in social science research.

After applying Rubin's Model to the three articles selected for analysis, ideas to clarify thoughts on the causal variable issue and to further the usefulness of this concept arose. These are 1) recognition that the causal interpretation of a variable depends on its relation to the unit of analysis (is the variable an attribute of the unit, or an attribute of something to which the unit is exposed); 2) the creation of a continuum of manipulation along which variables may be classified, thus eliminating the unnatural dichotomy of classification put forth by Holland and Rubin (unnatural

for social science variables); 3) the inattention to the issue of construct validity may contribute to the confusion about how to interpret results from demographic and personality variables; and 4) the use of proxy variables to measure the specific issues under study.

Unit of Analysis

One issue for deciding how to classify a variable is the relationship between the unit of analysis and the source of the variable. By source of variable I mean the origin of the variable values. The origin could emanate from the participant in the study (e.g., a participant's self rating of depression, a participant's race) or the origin could be completely independent of the participant (e.g., materials presented to the participant in a study, observer ratings of the participant). The *source* of the variable values matters for its interpretation as causal or not causal.

Social psychologists possess a history, a whole line of research, using the race and sex of experimenters as an independent causal variable, a line of work separate from studying the race or sex of participants. What would Rubin's model say about studies of experimenter effects, where the race, age, or sex of the experimenter is concluded as affecting responses of participants (e.g., Webb, Campbell, Schwartz & Sechrest, 1966)? Rubin's Model would render all that work obsolete.

In Article 2, three different sources exist for values of the variable 'gender.' For 'participant gender,' the values male and female originate with the participant, making it impossible for the participant to be exposed to multiple values of 'participant gender.' Gender, in that study was classified as both an 'S' and a 'G' variable. When 'harasser gender' is studied, values originate from materials created

by the researcher (vignettes and photos), making it possible for the participant to be exposed to multiple values of ‘harasser gender.’

The distinction Rubin does not make in his writing about the model is that, according to the potential exposability idea, attributes cannot be considered as causal when they are an attribute of the *unit under study*, the unit participating in the study. When the attribute is a controlled, *manipulated one*, such as the sex of person in a vignette as in this study, participants *can* be potentially exposed to a male harasser *or* a female harasser. And therefore conclusions about the attribute, ‘harasser gender,’ can be interpreted in a causal way.

The difference between values originating from the participants and materials created by the researcher matters for potential exposability. The study on sexual harassment demonstrates the possibility of manipulating an ‘attribute’ variable such as gender. In that study, when gender was manipulated through vignettes, participants were exposed to each level of the variable, male and female, and therefore, the pieces of the causal equation could be satisfied.

Another unit of analysis issue arose when analyzing the parenting and school readiness article. At first, the issue of what required manipulation arose. That is, did one need to be able to manipulate the parent’s behavior, or did one need to be able to manipulate the type of parenting behaviors to which a child was exposed? Rubin’s Model concerns the exposure of the unit of analysis to the various values of an independent variable. This distinction can become confusing when considering the application of these results. In the parenting article’s discussion section, the

applicability of the study results centers on a discussion of manipulating parental behaviors (parenting styles and involvement at home and school).

One could easily become confused about the source of manipulation when applying Rubin's Model to a study such as Article 1, where the data for the independent variable originates in one group of units (e.g., the parents) and the data for the dependent variables originate in a separate group of units (e.g., the children). Compare the following two examples: measuring a *child's* motivation as affecting *that child's* school performance versus a *parent's* style of behavior affecting her *child's* school performance. What becomes the 'thing' that the unit of study should be potentially exposable to? What has to be manipulable? What is the 'treatment'?

These issues raise an interesting point about classifying a causal variable according to the potential exposability idea. Two important issues arise: One from the standpoint of 'can the child be exposed to different levels of the variable being labeled as S'? The other from the standpoint of 'can the levels of 'S' actually be manipulated'? For RCM, the first question is the key question because in order for a valid comparison of conditions, the participant must be potentially exposable to more than one condition.

When the model stresses 'experimental manipulation' as its cornerstone (e.g., Holland & Rubin, 1983), it could lead to the misinterpretation and misapplication of manipulation in a study such as Article 1. Perhaps in social science studies, researchers should focus on the concept of exposing participants to various conditions of an independent variable, rather than on manipulating the independent variable. This would help researchers apply the model to studies in which the independent and

dependent measures originate from two separate groups of people as well as the more standard types of studies in which a researcher controls the independent variable, and in which both measures originate from the same group of people.

Continuum of manipulation

The issue of deciding what to classify as a causal versus not causal variable presents an immediate and crucial challenge to implementing Rubin's Model. Rubin and Holland state solid opinions about variables that can and cannot be considered causal. Holland writes, "...I take the position that causes are only those things that could, in principle, be treatments in experiments. The qualification 'in principle' is important because practical, ethical, and other considerations might make some experiments infeasible, that is, limit us to contemplating *hypothetical experiments*" (1986a, p. 954). "An attribute cannot be a cause in an experiment, because the notion of potential exposability does not apply to it. The only way for an attribute to change its value is for the unit to change in some way and no longer be the same unit" (1986a, p. 955).

Holland & Rubin articulate, "The 'experimental manipulation' is the focus of all causal inference in our model...The key notion is the potential for exposing each unit to any one of the experimental conditions in the study. For causal inference, it is critical that each unit be potentially exposable to any one of the experimental conditions" (Holland and Rubin, 1983, p. 6)

Variables in articles chosen for analysis in this dissertation present problems for the distinctions Holland and Rubin raise. Ethnicity and sex of participants (which includes the biological sex measures labeled as gender in Articles 2 and 3) clearly do

not meet the criteria spelled out above. But personality and physical attractiveness fit some of the criteria and not others. If a person's ethnicity or sex changed, fundamental aspects of the person would change as well. If a person's personality and attractiveness changed, some aspects of the person would change (e.g., attitudes and behaviors may change), but probably not to the same degree relevant to the change in ethnicity or sex. The distinctions among these variables' classifications as causal versus attribute variables would benefit from the creation of a *continuum of manipulation*.

With the idea of a continuum of manipulation, three areas of questioning must be considered. Note that this series of questions must be applied to the variables in each individual study, and not simply to a variable out of its study's context because as exemplified above, a single concept, such as 'biological sex' can be operationalized in several ways (e.g., as sex of research participants, sex of hypothetical persons in vignettes). These three areas of questioning are:

- 1) Is it possible to consider, regardless of practicality, that a person *could be exposed to*, or take on, another level of the variable under question? Before answering this question, consider the role that social construction and socialization play in many seemingly noncausal variables.

If no, then this variable is not at all manipulable and therefore not causal.

If yes, then:

- 2) Will anything else about the person change *in the process of*, (versus 'as a result of'), that exposure?

If no, then this variable is classified as clearly causal.

If yes, then:

- 3) What *level of change* in the person, as a result of the exposure, or taking on of this other level, is acceptable before it becomes difficult to assess change resulting solely from the actual variable of interest?

These *questions about exposing* participants to various levels of a variable create a *continuum of manipulation*, featured below, along which variables may be classified accordingly. On the left side of this continuum is the idea that if it is not possible to expose participants to different levels of a variable, regardless of practicality, then it is classified as not manipulable, and therefore not causal. Considering the outcome from changing such a variable would not offer much use in the way of applying results anyway. Moving to the right along the continuum the level of possible exposure and manipulation increases, as does the possibility that a variable can be classified as causal.

1	2	3	4	5
Not manipulable		Manipulable with careful attention to change in process		Manipulable
Not causal.....Causal				

Take, for example, the variable 'ethnicity' as used in the above analysis. Although at first glance the answer to question 1 (above) in relation to ethnicity would seem to be 'no,' the thought involved in making this decision becomes quite

complex. Ethnicity in Article 1 is defined as either Euro-American or African American. The ethnicity of the participant is the unit of focus. In answer to question 1, some might say it is not practical to consider changing a person's ethnicity. But this depends on how ethnicity is defined, a construct validity issue as will be discussed later.

We can imagine that it would be impossible to change a person's ethnicity when ethnicity is defined by biological factors (skin color, hair type, height, certain genetic diseases and predispositions) but ethnicity may also be defined by many social factors as well that a person has been exposed to because of socialization and discrimination, or even through customs and traditions.

Nancy Burson (Kelley, 2002) exemplifies a situation of manipulating race. She created a 'race machine' that photographs a person and produces pictures of how this person would look as a member of different racial groups. If such a machine was used in an experiment, participants of the study could be exposed to multiple races. But such a machine defines race through physical characteristics. Does this really indicate a change in race?

Clearly, the answer to the question of whether or not ethnicity and race are manipulable and causal is not an easy one to answer. Many social scientists measure ethnicity and race through self-reports of participants' belonging to ethnic and racial groups. Operationalizations of these variables based on biological and physical factors would fall at the lower end of the continuum of manipulation.

In contrast, in Article 3, the issue of personality as causal arises. Here, the idea of whether or not a person's personality can change without other aspects of the

person changing at the same time seems plausible, depending on which aspects of a person's personality change (i.e., shyness versus schizophrenia). Empirical research supports the understanding that nature and nurture shape personality in early childhood, and, as some support, through young adulthood (e.g., Costa & McCrae, 1988). Therefore, it is possible to imagine that one's personality could be different than what it is, to a certain extent. Personality therefore does not fall at the low end of the continuum, but would receive a rating around 2.

For another perspective on this continuum, take for example the variable 'gender' as used in Article 2. Going through the if/then statements above, can the participants be exposed to another level of the variable 'gender'? The researchers operationalize gender in a few different ways in that study, making it difficult to provide a yes/no answer to this question.

Returning to the unit of analysis discussion above, the answer to this questions is 'no' when gender (gender was the term used by the researchers, but it refers to biological sex in the context of that study) of the participants is under question, but 'yes' when the harasser and gender of the victim are in question. One could expose participants to a male harasser or a female harasser, or a male or female victim. Then, would anything else about the participants change in the process of being exposed to a male versus female harasser or male versus female victim? No. Therefore, according to the continuum of manipulation, but not according to RCM, gender of harasser and victim are clearly causal. Although 'participant gender' would receive a rating of '1' on the continuum, indicating classification as not causal, 'harasser

gender' and 'victim gender' as presented in the vignettes would receive ratings of '5' because they are manipulable.

Other variables, although not necessarily changeable during the course of a study, can be changed quite drastically. For example, the variable 'physical attractiveness,' used in Article 3, can drastically be changed. Physical attractiveness can be considered manipulable, because not only can a person choose how to present oneself, but they can also be physically altered, which these days is more and more common. Therefore, it is possible to expose an individual to different levels of physical attractiveness. Other things about the individual would change in the process, but not in such a way that the researcher would not be able to study the variable 'physical attractiveness' before and after exposure to various levels.

We can say for sure that attractiveness is manipulable to a degree, but we also do not know to what degree we would need to manipulate it in order to consider that a person's level of attractiveness is different than before, and we cannot be sure of all the variables that would change in the process, and as a result of changing attractiveness. Physical attractiveness therefore falls somewhere between a '3' and '4' on the scale.

What, exactly, determines a variable's place on this continuum? How much a change in the level of the variable connects to other aspects of the person (unit) changing? Or the degree of clarity in operationalizing the construct that can be ethically achieved? What defines manipulability? For Rubin, it would be the former – how much other aspects of the unit would have to change in order for the variable to be manipulated. But incorporating the latter, that construct validity must be

acknowledged as well in order to judge the manipulability of a variable, adds a necessary dimension to the model's decision process.

Construct Validity

Construct validity “involves making inferences from the sampling particulars of a study to the higher-order constructs they represent.” (Shadish et al., 2002).

Construct validity presents a major problem not easily addressed by RCM. RCM focuses attention on internal validity, ignoring, for the most part, construct validity (West, et al, 2000). RCM is also known as the ‘Potential-Response Model’ because of its focus on participants being potentially exposable to all values of a causal variable, and the structure of the causal effect equation. The Model's focus on responses, rather than causes of responses, situates it as a model concerned with internal, not construct, validity.

An important first step to understanding why construct validity should be considered as part of the model concerns the extant reasons supplied by proponents of RCM for why sex and race cannot be causal. The reasoning Rubin presents about not knowing exactly what about sex or race would cause an effect also applies to variables that seem clearly causal. Thinking about why RCM says we cannot consider race and sex to be causal raises problems with the ‘aspirin’ example Rubin has used (e.g., Little & Rubin, 2000; Rubin, 2002) to exemplify a causal variable. The analogous case of aspirin as treatment will be used to exemplify this point.

Rubin's example unfolds as follows: If one has a headache, one tries to decide whether or not to take an aspirin to relieve the headache. One considers what the headache would feel like after taking or not taking the aspirin. This sequence of

thinking follows the counterfactual approach to considering causal effects. RCM conceives the causal effect of the aspirin as the difference between potential outcomes in the future reduction of pain after taking and not taking the aspirin. The cause is taking aspirin. The assumptions are that there are not multiple versions of the treatment (all aspirins have the same effectiveness) and all units have the same reaction to the aspirin (Rubin, 2002).

Embedded in this example, however, are assumptions about the causal mechanism, and, ultimately, the cause itself. One assumption implies that the cause of reduction in pain is the effectiveness of the aspirin. But is the cause the *effectiveness of the chemical* in the aspirin? Or the *effectiveness of the mere action* of taking a pill? Plenty of research supports the effect of taking a placebo.

While a study with two groups – those taking aspirin and those taking no aspirin – might show a causal effect of aspirin, a study with two groups – one taking aspirin and the other taking placebo – might show no causal effect. The same cause may be acting in both of these studies, or two different causes may be acting, one in each study. The idea being that the aspirin itself includes many elements, such as the idea of taking a pill. Does the cause change depending on what we compare it to? Yes, it does. The causal effect depends on the levels chosen for comparison. Also, what changes is the level of specification of what the cause is. This becomes more important when we change what we are comparing it to. How we define the “construct” is the key to specifying the cause. Just as we may not know exactly what about ‘sex’ or ‘race’ ‘causes’ an outcome, we also may not know exactly what about ‘aspirin’ causes an outcome.

Considering construct validity also leads us to question how it is that sex (or gender as labeled in the article) in Article 2 can be causal in one sense (i.e., when manipulated in the vignette) but not causal in another sense (i.e., when an attribute of participants). The difference between these two forms of operationalizing ‘gender’ pertains to a construct validity issue. The contradiction between ‘gender’ being causal in one operationalization but not another results from ‘gender’ not being the ultimate cause.

An issue complicating the use of the manipulation of sex while keeping all other aspects of the conditions constant is the fact that photographs were also shown to participants to manipulate the gender of the persons in the vignettes. The photographs are supposed to be photos of the persons discussed in the vignettes. Although participants can be potentially exposed to the various photographs, other things about the persons in these pictures differ among pictures, making it hard to say that the photos are only manipulating sex. The researchers did consider confounding variables and controlled for the physical attractiveness and age of persons in these photos.

Although the researchers focus on gender of person in photo as their variable under manipulation, it is not clear that what the participants are reacting to in the photos is the biological gender of the person. More likely it is the gender stereotype or gender role, or power or social norms, all of which gender is a possible proxy for. In the actual vignettes, issues such as homophobia may have been at work as a possible proxy as well.

This is a construct validity issue, an issue RCM does not touch upon. It confuses the issue of what exactly the researcher was manipulating and what exactly the participants were responding to. In this sense, Rubin's 'constancy effect' assumption may be in jeopardy. That is, if different participants are paying attention to different aspects of the manipulation materials, the causal effect may *not* be constant across participants.

Although these harasser and victim variables are used to create the conditions in Article 2, are they really 'S' variables? The way they are presented, they are S variables. What makes it possible for these to be presented as manipulable variables is that it is not the gender of an actual person that is being manipulated - that could not be possible. Instead, in this study, 'equivalent' photos, (equivalent in terms of age and physical attractiveness, to rule these two variables out as alternative rival hypotheses) of men and women were presented to participants, with the idea in mind that all that was physically different about the pictures was the biological sex of the people in the photos, and the names in the written materials supplied to participants. So the assumption here is that all the participants perceive as being different in these written documents and photos is biological sex.

An additional issue to consider in applying these results does not mesh with RCM: If a person approached a manager with a similar case, and the manager was to make a decision about whether or not a person was mistreated because of her biological sex given Rubin's Model, how would he proceed then?

Social scientists have been dealing with these issues of construct validity. The article, "Social science research on trial: Use of sex stereotyping research in Price

Waterhouse v. Hopkins” (Fiske et al, 1991), which presents a sex discrimination case in which social scientists were called to provide expert testimony, provides a perfect example. “The Judge ultimately ruled ‘an employer that treats [a] woman with [an] assertive personality in a different manner than if she had been a man is guilty of sex discrimination’” (Fiske et al, 1991, p. 1052). The issue of whether or not the woman bringing harassment charges was treated unfairly because of her personality or her biological sex was a prominent one in the trial: “We sit not to determine whether Ms. Hopkins is nice, but to decide whether the partners reacted negatively to her personality because she is a woman” (p. 1054).

Proxy Variables

The causal variable definition presents such a problem for social science research partly because it neglects to incorporate the idea of construct validity. But as the discussion here about construct validity highlights, if sex and race are used to represent so many other variables, why not just test these other variables instead of relying so heavily on simple sex and race category measures? Explanations of Rubin’s Model deem race and sex as not causal variables because other aspects of a person would have to change in order for the race and sex of the unit to change, and because we do not know what exactly about sex or race causes an effect. A logical next step to address this causality problem for sex and race entails further contemplation about how to handle these variables in the causal realm, rather than simply removing them from that realm altogether, as Rubin’s Model does.

RCM dictates that an attribute independent variable results in descriptive conclusions. As has been mentioned repeatedly throughout this paper, researchers,

policy makers, expert witnesses and juries likely do not place the same weight and importance on descriptive conclusions as they do on causal conclusions. How are descriptive conclusions applied? What does one do after discovering a descriptive relationship? Perhaps the Elaboration Model would say that such a relationship lends itself to a test variable to begin elaborating the relationship.

Rather than simply categorizing attribute variables as producing descriptive conclusions, and risking the possible consequence of such conclusions being overlooked, a second solution will be proposed: the use of proxy variables. Proxy variables offer a possible solution for bridging the gap between attributes not being causal variables, and the need for social scientists to study these variables. The idea of a proxy variable becomes relevant here to acknowledge the leap between what is being measured and what is *intended* to be measured. With the variable sex, beyond the potential exposability problem, there is confusion about what about sex causes a difference (i.e., a construct validity problem). Is it biological sex, sex stereotypes, social roles? In a similar vein, the concept of race represents and is represented by myriad variables. The concept of categorizing people by race changes over time, demonstrated the socially constructed nature of this variable, as well (American Sociological Association, 2002).

While acknowledging the existence of proxy variables is a necessary first step to addressing the attributes as non-causal variable issue, this is not enough. How many times does a researcher collect data about biological sex when in fact the concept being studied is one of gender socialization, gender norms, gender roles, mostly measurable social constructions grounded in the biological sex dichotomy? By

following RCM's causal variable definition, we prevent biological sex from being labeled as a cause of gender differences. While it may be useful to identify biological sex differences in areas such as medicine, in the social sciences, as well as in some areas of medicine (e.g., healthcare), the implications of sex differences can be socially detrimental. But at the same time we overlook the implications, that what researchers in the social sciences often measure as biological sex really falls under the gender classification, and most concepts of gender, some as listed above, *are* manipulable.

This is not a new idea – that sex and gender often become confused in the literature. However, it continues to be problematic and the application of RCM highlights the danger in this confusion.

Organizations such as APA require researchers to report studies in a particular way; they require certain research guidelines and ethical practices be followed. To advance the validity of conclusions about gender and race findings, researchers should be required to include alternative measures of race/gender when possible, beyond the 'are you male or female' measure, in order to publish race or gender difference results.

For example, a researcher investigating gender differences in sexual harassment as in Article 2 could also include, or instead include, measures of benevolent/hostile sexism (Glick & Fiske, 1996), sex-roles (Bem, 1974), gender norms, power, or gender stereotypes to assess how participants differ in ratings of male/female harassers/victims. These concepts *are manipulable* and would therefore provide causal variables and causal inferences. This would also address Sherif's

(1994) concern about the black-box idea of what we are measuring when we measure sex (or race, or any attribute).

It may not be as possible to include such additional measures when the sex or race of research participants is under study, as opposed to the sex or race of a person in a vignette or an experimenter. But sex-role measures or beliefs about gender norms measures filled out by participants may provide gendered information about how participants think and behave. Such supporting information and measures should be required for research. Otherwise the field will go on criticizing itself and its researchers for improperly studying sex and race, and the conclusions will continue to exist in a liminal space.

If these seem like simple ideas, or practices researchers should already consider, then why does the field not hold researchers to a higher standard on these issues? Why do researchers not already follow these practices? In some disciplines, such as women's studies, researchers do follow such practices. As the former editor of *Sex Roles: A Journal of Research*, Sue Rosenberg Zalk, commented (and actively tried to change) many research articles about gender differences result from the mere inclusion of a biological sex measure (man/woman) that resulted in statistical significance when nothing else in the study did. She believed that in order to publish meaningful sex role or gender research, sex or gender should be part of a study from its inception so that relevant theoretical considerations would be involved in the study's design and the appropriate measures and variables would be included to draw informative conclusions.

Limitations

As in any study, the data collected may limit the issues that arose as problematic and not problematic. For this methodological study, the choice of research articles used as data was not meant to produce a random sample of articles or a representative sample of articles. Methodological articles do not focus on random samples of examples as much as examples that fit the questions under study. It is possible that many articles from various social science disciplines could have offered interesting challenges to the Rubin's Causal Model. The articles chosen, though, did raise critical issues and challenges for RCM. And the conclusions drawn in this dissertation can be generalized to other studies, regardless of discipline, that share the problems and issues addressed here.

Classifying language as causal versus non-causal became difficult because 1) researchers did not use language consistently, and 2) because no pre-established causal versus non-causal language categorization list exists. Also, with terms such as 'moderate,' the analysis of causal inferences relied somewhat on conventions of interpretation for the process (e.g., Baron & Kenny, 1986). The difficulty classifying language impacted what was and was not classified as a causal inference.

The focus of this dissertation was one of methodology, not statistics. However, had I possessed the data used in the three articles, I could have reanalyzed that data to derive more details for applying the Elaboration Model. As it was, applying the EM proved difficult in Articles 1 and 2 because the logic of the articles did not mesh with the logic of the EM. Analyzing the data from those articles would allow me to follow the EM's process more thoroughly than was possible here.

Guidelines

Researchers would benefit from approaching research with all three of these methodological models in mind to evaluate the logic and methods of reaching causal conclusions. The following guidelines resulting from this dissertation provide ideas researchers may wish to consider when pursuing causal research in the social sciences.

I. In the design stages:

A. Apply basic tenets of the Elaboration Model

- i. Consider theoretically relevant variables so that test variables may be measured and included in the analysis
- ii. Consider the temporal sequence of variables and the implication of this sequence for the conclusions

B. Consider the Validity Typology

- i. Translate the study design into the Validity Typology's notation to highlight the design's structure and related plausible threats to validity
- ii. Consider plausible alternative hypotheses in the realms of all four validities so that measures may be taken to rule out plausible threats in advance

C. Think about the framework of Rubin's Causal Model

- i. What is the unit of analysis? Keep the unit of analysis in mind when classifying variables according to RCM notation and answering questions about potential exposability

ii. Classify each variable according to Rubin's Causal Model's

notation: S, G, X, Y

1. Does the independent variable fit the definition of a causal variable (an S variable)?
 - a. Answer the questions relevant to the continuum of manipulation:
 - i. Is it possible to consider the person exposed to another level of the variable?
 - ii. Will anything else about the person change in that process of exposure?
 - iii. What level of change is acceptable before it becomes difficult to distinguish the desired change from other changes occurring in the process?
2. Does the independent variable fit the definition of an attribute variable (a G or an A variable)?
 - a. Is this variable a proxy variable?
 - i. If yes, can the variable that the attribute acts as a proxy for be included in the study?
 - ii. If not, why not?
3. Are there any covariates, or X, variables?

4. What will the possible values of Y be? Consider the answer to this question according to how variables were classified as S or G above
5. Construct the causal effect equation, considering
 - a. How variables have been classified with RCM notation
 - b. What pieces are missing for an analysis of causal effects?

II. In the analysis stage:

A. Following the Elaboration Model

- i. If a relationship at first demonstrates no statistical significance, include other measured variables (already in the study as test variables) to identify a possible suppressor or distorter relationship
- ii. If a partial relationship results in replication, this should also receive attention because it, too, offers information to further theory
- iii. Insure that the order and relationships tested in the statistical analyses match the order and relationships originally proposed in hypotheses

B. Using Rubin's Causal Model

- i. Check to insure that the comparisons defined by the causal effect equation are the comparisons being tested in the statistical analyses

III. In the inference stage:

A. Consider the Elaboration Model

- i. Use the EM's classification of partial relationships to categorize the relationship resulting from the addition of test variables
 - 1. First recall the temporal sequence of the test variable in relation to the main variables of interest
 - 2. Is the resulting partial relationship explanation, interpretation, specification, replication, distortion, suppression?
 - 3. Are the interpretations of partial relationships matching the results of the statistical analyses and the variables put proposed in the hypotheses?

B. Use the Validity Typology

- i. Revisit plausible alternative hypotheses and determine the degree to which they can be ruled out and the degree to which they jeopardize the validity of causal conclusions

C. Consider the perspective of Rubin's Causal Model

- i. Use the causal effect equations formed above as a guide to drawing inferences

1. Address the specific comparisons as spelled out in the equations formed above
 2. Note any conditional relations and ensure that these are clearly expressed
- ii. Revisit the classification of variables according to notation and the continuum of manipulation and abide by the classification decisions in the inferring of conclusions (S, G)
 - iii. Do the results from the equation match the classification of variables as causal vs. non-causal?
 - iv. How are interactions being interpreted?
 - v. If proxy variables were not used, are attribute variables being interpreted as explanations or proxies?

CHAPTER 14: CONCLUSION

The language of causation begets a high level of responsibility for those who use it. The idea of causation encompasses a sense of certainty and permanence. When a researcher concludes, for example, that race determines a child's performance in school, the connection, once communicated to colleagues and the public, takes on a dangerous political permanency that bends, but may not easily break, in the presence of competing ideas.

Avoiding Misinterpretation and Misuse

The scientific implications carried by conclusions are not separate from the moral or ethical implications of those conclusions. Drawing a conclusion that "race causes" indicates an internal attribution. Such a conclusion places responsibility (i.e., cause) inside the person and removes any responsibility (i.e., cause) from the environment, from institutions, and from social structures. Such a conclusion also serves a political purpose, promoting social stratification, discrimination, and other types of inequality.

The Language of Validity

Given the power that conclusions carry, researchers need to be much more explicit about what they are saying in a conclusion. Leaving the reader to interpret whether or not a conclusion might be causal and which aspects of the research enabled the researcher to draw causal conclusions opens a dangerous door to misinterpretation and misuse. How, then, can a researcher mitigate misinterpretation and misuse?

Conscious Language Choices

First, researchers can make conscious language choices, clearly communicating to the reader whether or not an inference is causal, predictive, or associative. Clear language choices include being consistent throughout a research paper, and not alternating between causative, predictive and associative language when referring to a relationship between variables. The language used in the title, abstract, introduction, hypotheses, methodology, results, and discussion should exemplify continuity.

Explicitly Acknowledge and Discuss Causality

Second, researchers can explicitly discuss the issue of causality, beyond their specific word choices, throughout a research paper. Social science textbooks discuss quasi-experimental and nonexperimental designs as ‘alternatives’ to experiments, when ethical and moral implications prevent processes such as random assignment from taking place, thereby clouding the process of drawing clear causal inferences. This indicates a belief in the field that being able to draw causal inferences remains a primary goal. Studies not including ‘normative’ processes such as random assignment may dance around the issue of cause, wishing to make causal inferences, yet being careful not to make unwarranted causal statements, and ultimately creating an undertone of causality throughout a research article.

Therefore, whether the relationships under study fall in the causal realm or not, a discussion of causality may still be highly relevant in an article. This discussion of causality can consist of the myriad concepts discussed throughout this dissertation. There is no shortage of approaches to studying causal relationships.

Researchers can use the approach of the Elaboration Model for an explicit discussion about additional variables considered important to the relationships under study in a paper. Whether or not a resulting partial relationship demonstrates significance, the additional knowledge the researcher gains by adding a test variable advances theoretical understanding of the relationships under study. The Elaboration Model encourages researchers to acknowledge nonsignificant relationships, adding to the thoroughness of an investigation and the advancement of theory.

The Elaboration Model also offers researchers an additional vocabulary consisting of the nouns ‘specification,’ ‘replication,’ ‘explanation,’ ‘interpretation,’ ‘suppression’ and ‘distortion’ to classify relationships. A researcher implementing the Elaboration Model gains the ability to provide more in-depth explanations of relationships, a clear step towards increasing the validity of a causal inference.

Researchers can use the Validity Typology to guide the choice of structural elements in the research process. Taking time to consider plausible rival hypotheses before carrying out a study will increase a researcher’s confidence in drawing causal inferences. While the Elaboration Model encourages the researcher to consider what additional test variables might be included, the Validity Typology encourages the researcher to engage in a sort of opposite thought process – what variables can be eliminated from a study. Both types of thinking result in a more thorough investigation and increase the validity in causal inferences.

A researcher applying the Validity Typology will consider maturation, attrition and other inconspicuous processes intricately tied to the validity of a causal inference. Not overlapping the Elaboration Model’s focus on theoretical depth, the

Validity Typology offers the researcher a framework of questions guiding the construction of a research study, allowing the researcher to increase the validity in resulting causal inferences from the beginning of the research process.

Researchers would benefit from learning and using Rubin's Causal Model. Rubin's Model asks the researcher to approach the study of a relationship from the effects being studied, rather than the causes being studied. This leads the researcher down a path of consciously considering the prospect of potential responses, as spelled out in Rubin's Model's causal effect equation.

The process of attempting to apply the causal effect equation, will, in itself, raise the researcher's awareness of assumptions inherent in the study design, variable value choices, and possible conclusions. Categorizing the variables according to the symbolic language of the model will also engage the researcher in a thought process about the types of variables in the study and their respective values. In addition, the analysis of Rubin's Model in this dissertation sheds light on new aspects of the unit of analysis issue, possible additional notation for the model, and consideration of the match between types of variables chosen and the types of conclusions they might warrant.

Rubin's Model's causal variable definition highlights the need for researchers to consciously acknowledge whether an independent variable is considered causal or descriptive. The model offers one set of guidelines for this decision process. To aid in this process, this dissertation has also proposed the continuum of manipulation to guide the decision process of whether or not a variable might be considered causal. This dissertation has also focused on the idea of proxy variables, and the need for a

researcher's choice of variables and operationalizations to occur at the deepest level possible for capturing the actual concept under study, thereby leading to an increase in the validity of causal inferences.

Summary

Research studies raise more questions than they answer, yet we are taught to conclude published research articles with a discussion of what we 'know' as a result of having conducted the study, including only a few paragraphs on limitations. The purpose of conducting research is to advance our understanding and ultimately our knowledge. A researcher engages in knowledge advancement through process of drawing conclusions and communicating these conclusions.

A researcher should develop an ongoing dialogue about causation throughout a research article. This dialogue should address the steps taken to increase the validity of causal inferences, steps from the inception of a research study through the process of inferring a conclusion. This dialogue would best consist of the specific concepts and ideas outlined in detail in this dissertation: the Elaboration Model, the Validity Typology and Rubin's Causal Model, as well as the concepts resulting from critiques of these models. A researcher who takes responsibility for explicitly evaluating the validity of inferences in her own study will greatly increase the probability of a researcher, policy maker or community activist accurately applying her conclusions, thereby truly contributing to the application and advancement of social science.

Table 1

Notation for Rubin's Causal Model

Symbol	Meaning	Example
S	Exposure variable	S = t indicates a participants has been exposed to the treatment condition of a study, whereas S = c indicates a participant has been exposed to the control condition
G	Subgrouping variable	G = m indicates a participant is a man, whereas G = w indicates a participant is a woman
A	Attribute variable	Note: This is Holland's notation, and not Rubin's. A = a, where 'a' is the value of a participant on an attribute
X	Covariate	X = pretest score, and this score should not differ for a participant depending upon the condition of a study to which she is exposed

Table 1 Cont.

Y	Outcome variable	Y_t is the outcome (Y) for a participant exposed to the treatment condition (t)
---	------------------	---

Table 2

Example of a Randomized Study

Participant	S	Treatment condition	Control condition
Joseph	T	3	5
Carol	T	4	6
Jennifer	C	3	5
Steven	C	4	6
Mean response		3.5	5.5

Note: Numbers in bold represent observed values.

Table 3

Example of a Randomized Study Using Rubin's Causal Model Notation

Participant	Treatment condition	Control condition	$Y_t - Y_c$
Joseph	$Y_t = 3$	$Y_c = 5$	-2
Carol	$Y_t = 4$	$Y_c = 6$	-2
Jennifer	$Y_t = 3$	$Y_c = 5$	-2
Steven	$Y_t = 4$	$Y_c = 6$	-2
Mean response	$E(Y_t) = 3.5$	$E(Y_c) = 5.5$	$E(Y_t - Y_c) = -2$

Note: Numbers in bold represent observed values.

Table 4

Example of a Nonrandomized Study Using Rubin's Causal Model Notation

Participant	Treatment condition	Control condition	$Y_t - Y_c$
Joseph	$Y_t = 3$	$Y_c = 5$	-2
Carol	$Y_t = 4$	$Y_c = \mathbf{6}$	-2
Jennifer	$Y_t = 3$	$Y_c = 5$	-2
Steven	$Y_t = 4$	$Y_c = \mathbf{6}$	-2
Mean response	$E(Y_t) = 3.5$	$E(Y_c) = 5.5$	$E(Y_t - Y_c) = -2$

Note: Numbers in bold represent observed values.

Table 5

Frequencies for Bivariate Relationship between Religion and Age

DV: Is religion important to you		
IV: Age	Yes	No
Young	10	40
Old	40	10

Table 6

Partial Relationships Formed from the Test Variable 'Sex'

Test variable: Sex				
Men		Women		
DV: Religion important				
Age	Yes	No	Yes	No
Young	5	20	5	20
Old	20	5	20	5

Table 7

Partial Relationships Classification for the Elaboration Model

	Temporal occurrence of test variable	
	Antecedent	Intervening
Partial relationship compared with original		
Same relationship	Replication	Replication
Less or none	Explanation	Interpretation
Split	Specification	Specification

Note: This table reprinted from *The Practice of Social Research* (Babbie, 2001).

Table 8

Examples of Causal Language

Obvious verbs	Not so obvious verbs	Nouns	Other
affects	accounts for	consequence	why
causes	based on	determinant	because
determines	buffers	effect	
influences	creates	evidence of	
produces	depends on	function of	
	diminishes	mechanism	
	due to	origin of	
	enhances	process	
	facilitates		
	hinders		
	improves		
	leads to		
	makes		
	moderates		
	plays a role in		
	promotes		
	reduces		
	results in		
	shapes		
	stimulates		

Table 9

Examples of Associational, Predictive and Ambiguous language

Associational	Predictive	Ambiguous
associated with	predict	benefits from
correlates with	guides	explains
covaries with		interferes
links to		pathway to
relates to		the role of
supports		

Table 10

Bivariate Relationship between Harasser and Verdict

		Verdict	
Harasser		Guilty	Not-guilty
Male harasser			
Female harasser			

Table 11

The Role of Gender Combination in Relation to Harasser Gender and Verdict

Harasser	Cross-gender		Same-gender	
	Guilty	Not-Guilty	Guilty	Not-Guilty
Male harasser				
Female harasser				

Table 12

Summary of Comparisons among the Three Models After Applications to Articles

Issue	RCM	EM	VT
Variables			
<i>Nature/Type</i>			
1. Attributes are problematic causes	Y*	N	N
2. Considers manipulability of independent variables	Y*	N	N
3. Leads to categorizing variables (IVs, subgrouping, test variables)	Y*	Y	N
<i>Constructs</i>			
4. Considers construct validity	N*	Y	Y
5. Operationalization matters	Y*	N	Y
Focus			
6. Focuses on causes of effects	Y	Y	Y
7. Focuses on effects of causes	Y	Y	N
Order			
8. Considers temporal order of variables	Y	Y	Y
9. Considers order in which relationships were tested	N	Y	N
10. Considers relationship between	N	Y	N

Table 12 Cont.

hypotheses and analyses run			
	Explanation		
11. Considers mechanism	N	Y	N
12. Considers plausible rival	N	Y	Y
hypotheses			
	Research design/analysis		
13. Selection of participants matters	Y	N	Y
14. Unit of analysis matters	Y*	N	N
15. Takes into account research design	N	N	Y
16. Considers procedure of research	N	N	Y
17. Type of data matters (cont vs cat)	Y*	Y	N
18. Considers significant as well as	N	Y	N
nonsignificant results			
Note: N indicates 'no, the model does not consider this issue' and Y indicates 'yes, the model does consider this issue'; * represents an issue expanded upon in the discussion section.			

References

- American Sociological Association. (2002, August). Executive summary statement of the American Sociological Association on the importance of collecting data and doing social scientific research on race. Washington DC: ASA.
- Azar, B. (2002, September). Searching for genes that explain our personalities. Monitor on psychology: A publication of the American Psychological Association, 33 (8) 44-46.
- Babbie, E. (1992). The elaboration model. In E. Babbie, The practice of social research (pp. 409-428). Belmont: Wadsworth.
- Babbie, E. (2001). The practice of social research. Belmont: Wadsworth/Thomson Learning.
- Bem, D. J. (1987). Writing the empirical journal article. In M. P. Zanna & J. M. Darley (Eds.), The compleat academic (pp. 171-201). New York: McGraw-Hill.
- Bem, S. L. (1974). The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 42, 155-162.
- Blalock, H.M. (1961a). Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina.
- Blalock, H.M. (Dec., 1961b). Evaluating the relative importance of variables. American Sociological Review, 26 (6), 866-874.
- Boruch, R. F., & Rindskopf, D. (1977). On randomized experiments, approximations to experiments, and data analysis. In L. Rutman (Ed.), Evaluation research methods: A basic guide (pp. 143-176). Beverly Hills: Sage.

Buss, D. M. & Schmitt, D. P. (1993). Sexual strategies: An evolutionary perspective on human mating. Psychological Review, 100, 204-232.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company.

Carrington, P. & Dembski, J. (2002). *Lecture eight: Multivariate data analysis*. Retrieved September 20, 2002, from http://www.arts.uwaterloo.ca/SOC/courses/soc321/Lecture_8.html.

Cochran, W.G. (1965). The planning of observational studies of human populations. Journal of the Royal Statistical Society, Series A (General), 128 (2), 234-266.

Cochran, W.G. (1971). Analysis of covariance: Its nature and uses. In J.A. Steger, (Ed.), Readings in statistics for the behavioral scientist (pp. 179-193). New York: Holt, Rinehart, and Winston, Inc.

Cochran, W.G. (1983). Planning and analysis of observational studies. New York: John Wiley & Sons.

Cochran & Rubin (1973). Controlling bias in observational studies: A review. Sankhya, Series A, 35 (4), 417-446.

Cohen, J. (1992). Fuzzy Methodology. Psychological Bulletin, 112 (3), 409-410.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum Associates.

Collins, P. H. (1991). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. New York: Routledge.

Cook, T. D. & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton-Mifflin.

Cook, T. D. & Campbell, D. T. (1983). The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 223-326). New York: John Wiley & Sons.

Costa, P.T., & McCrae, R. R. (1988). Personality in Adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology, 54(5), 853-863.

Cox, D. R. (1986). Comment. Journal of the American Statistical Association, 81, 963-964.

Cox, D. R. (1992). Causality: Some statistical aspects. Journal of the Royal Statistical Society, Series A, 155 Part 2, 291-301.

Dawid, A. P. (2000). Causal inference without counterfactuals. Journal of the American Statistical Association, 95 (450), 407-448.

Deaux, K. (1993). Commentary: Sorry, wrong number – a reply to Gentile's call. Psychological Science, 4 (2), 125-126.

Denmark, F., Rabinowitz, V., & Sechzer, J. (2000). Engendering psychology: Bringing women into focus. Boston: Allyn and Bacon.

DeVault, M. L. (1996). Talking back to sociology: Distinctive contributions of feminist methodology. Annual Review of Sociology, 22, 29-50.

Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is what is good. Journal of Personality and Social Psychology, 24, 285-290.

Dohrenwend, B. P., Levav, I., Shrout, P.E., Schwartz, S., Naveh, G., Link, B.G., Skodol, A. E., & Steuve, A. (1992). Socioeconomic status and psychiatric disorders: The causation-selection issue. Science, 225, 946-952.

Eagly, A. H. (1995). The science and politics of comparing women and men. American Psychologist, 50 (3), 145-158.

Eisenman, R. (1995). Why psychologists should study race. American Psychologist, 50 (1), 42-43.

Evans, S. H. & Anastasio, E. J.. (1968). Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 69 (4), 225- 234.

Fearon, J.D. (1991). Counterfactuals and hypothesis testing in political science. World Politics, 43 (2), 169-195.

Ferber, R. (1966). Item nonresponse in a consumer survey. Public Opinion Quarterly, 35, 275-290.

Fine, M. (1996). "Whiting out" social justice. In R.T. Carter (Ed.), Addressing cultural issues in organizations: Beyond the corporate context (pp. 35-50). London: Sage.

Fine, M. & Barreras, R. (2001). To be of use... Analyses of Social Issues and Public Policy, 1 (1).

Fiske, S. T., Bersoff, D. N., Borgida, E., Deax, K., Heilman, M. E. (1991). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins. American Psychologist, 46 (10), 1049-1060.

Fivush, R., Brotman, M. A., Buckner, J. P., & Goodman, S. H. (2000). Gender difference in parent-child emotion narratives. Sex Roles: A Journal of Research, 42 (¾), 233-253.

- Francis, J. & Busch, L. (1975). What we don't know about 'I don't knows.' Public Opinion Quarterly, 39, 207-218.
- Gagnon, S. (1996, June). Cause and consequence markers in scientific popularizing work: Traces of the explicative paradigm. Paper presented at the 1st International multidisciplinary conference: Knowledge & discourse: Changing relationships across academic disciplines and professional practices, Hong Kong.
- Ghiselli, E. E., Campbell, J. P., Zedeck, S. (1981). Measurement theory for the behavioral sciences. San Francisco: W.H. Freeman.
- Glass, G.V. & Hopkins, K.D. (1996). Statistical methods in education and psychology (3rd ed.). Boston: Allyn & Bacon.
- Glick, P, & Fiske, S. T, (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. Journal of Personality and Social Psychology, 70, 491-512.
- Glymour, C. (1986). Comment: Statistics and metaphysics. Journal of the American Statistical Association, 81 (396), 964-966.
- Holland, P.W. (1986a). Statistics and causal inference. Journal of the American Statistical Association, 81 (396), 945-960.
- Holland, P.W. (1986b, Sept-Oct.). Which comes first, cause or effect? The New York Statistician, 38 (1), 1-2,4,6.
- Holland, P.W. (1986c). Rejoinder. Journal of the American Statistical Association, 81 (396), 968-970.

Holland, P.W. & Rubin, D.R. (1983). On Lord's Paradox. In H. Wainer & S. Messick (Eds.), Principals of modern psychological measurement: A festschrift for Frederic M. Lord (pp. 3-25). New Jersey: Laurence Erlbaum Associates.

Jasanoff, S. (2001, May/June). Judicial fictions: The Supreme Court's quest for good science. Society, 27-36.

Jones, J. (1997). Whites are from Mars, O.J. is from Planet Hollywood: Blacks don't support O.J. and Whites just don't get it. In M. Fine, L. Weis, L.C. Powell, & L.M. Wong (Eds.), Off White: Readings on race, power and society (pp.251-258). New York: Routledge.

Joreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), Applications of statistics (pp.265-287). Amsterdam: North-Holland.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). Research methods in social relations. New York: Harcourt Brace Jovanovich College Publishers.

Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. Psychological Bulletin, 64 (5), 326-329.

Kaufman, J. S. & Cooper, R. S. (1999). Seeking causal explanations in social epidemiology. American Journal of Epidemiology, 150 (2), 113-120.

Kaufman, J. S., Cooper, R. S., & McGee, D. L. (1997). Socioeconomic status and health in Blacks and Whites: The problem of residual confounding and the resiliency of race. Epidemiology, 8 (6), 621-628.

Kelley, T. (2002, April 14). Through machine, seeing more of others in yourself. The New York Times, p. 37.

Kenrick, D.T., Sundie, J. M., Nicastle, L. D. & Stone, G. O. (2001). Can one ever be too wealthy or too chaste? Searching for nonlinearities in mate judgement. Journal of Personality and Social Psychology, 80 (3), 462-471.

Laner, M.R. (2000). "Sex" versus "Gender": A renewed plea. Sociological Inquiry, 70 (4), 462-474.

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Cambridge: Harvard University Press.

LaVeist, T.A. (1994). Beyond dummy variables and sample selection: What health services researchers ought to know about race as a variable. Health Services Research, 29 (1), 1-16.

Lazarsfeld, P. F. (1955). "Foreword," In Hyman, H., Survey design and analysis: Principles, cases and procedures (pp. ix-xvii). Illinois: The Free Press.

Lazarsfeld, P. F., Pasanella, A. K. & Rosenberg, M. (Eds.). (1972). Continuities in the language of social research. New York: The Free Press.

Leamer, E. E. (1988). Discussion. In C. C. Clogg (Ed.), Sociological methodology (pp. 485-493). Washington D.C.: American Sociological Association.

Licht, M. H. (2000). Multiple regression and correlation. In L.G. Grimm & P.R. Yarnold (Eds.) Reading and understanding multivariate statistics (pp. 19-64). Washington D.C.: American Psychological Association.

Lindsey, E. W., Mize, J. & Pettit, G. S. (1997). Differential play patterns of mothers and fathers of sons and daughters: Implications for children's gender role development. Sex Roles: A Journal of Research, 37 (9/10), 643-661.

Little, R. J. & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. Annual Review of Public Health, 21, 121-45.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 68 (5), 304-305.

Lord, F.M. (1969). Statistical adjustments when comparing preexisting groups. Psychological Bulletin, 72 (5), 336-337.

Marini, M. M., & Singer, B. (1988). Causality in the social sciences. In C. C. Clogg (Ed.), Sociological methodology (pp. 347-410). Washington D.C.: American Sociological Association.

Matthews, K.A. (1989). Are sociodemographic variables markers for psychological determinants of health? Health Psychology, 8 (6), 641-648.

Maxwell, S. E., & Cole, D. A. (1995). Tips for writing (and reading) methodological articles. Psychological Bulletin, 118 (2), 193-198.

McIntosh, P. (1989, July/August). White privilege: Unpacking the invisible knapsack. Peace and Freedom, 10-12.

Meyerowitz, B. E., Chaiken, S. & Clark, L. K. (1988). Sex roles and culture: Social and personal reactions to breast cancer. In M. Fine & A. Asch (Eds.), Women with disabilities: Essays in psychology culture and politics (pp. 72-89). Philadelphia: Temple University Press.

Moore, D.S. (1997). Statistics: Concepts and controversies. W.H. Freeman and Co.: New York.

Narayan, S. & Krosnick, J. (1996). Education moderates some response effects in attitude measurement. Public Opinion Quarterly, 60, 58-88.

Neyman, J. S. (1990). On the application of probability theory to agricultural experiments: Essay on principles, Section 9. Dabrowska, D.M. & Speed, T.P. (Eds. and Trans.), Statistical Science, 5 4, 465-480. (Reprinted from Annals of Agricultural Sciences, 1-51, 1923.)

Parker, I. (1997). Discursive psychology in Fox, D & Prilleltensky, I (Eds.) Critical psychology: An introduction. London: Sage

Pearl, J. (1997, February). The new challenge: From a century of statistics to an age of causation. Technical report (R-249). Presentation at the IASC Second World Congress, Pasadena, CA.

Pepitone, A. (2000). A social psychology perspective on the study of culture: An eye on the road to interdisciplinarianism. Cross-Cultural Research, 34 (3), 233-249.

Pinderhughes, H. & Shim, J. K. (2001). The social production of racial inequalities in health. Unpublished manuscript, University of California, San Francisco.

Phares, J. E. (1991). Introduction to personality (3rd ed.). New York: HarperCollins.

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? American Psychologist, 51 (9), 918-927.

Potter, J. & Wetherall, M. (1987). Discourse and social psychology: Beyond attitudes and behavior. London: Sage.

Publication manual of the American Psychological Association (4th ed.). (1994). Washington DC: American Psychological Association.

Publication manual of the American Psychological Association (5th ed.). (2001).

Washington DC: American Psychological Association.

Raag, T. (1999). Influences of social expectations of gender, gender stereotypes, and situational constraints on children's toy choices. Sex Roles: A Journal of Research, 41 (11/12), 809-832.

Rein, M. & Winship, C. (1999, July/August). The dangers of 'strong' causal reasoning in social policy. Society, 36 (5), 38-47.

Rindskopf, D.M. (1978). Secondary analysis: Using multiple analysis approaches with Head Start and Title I data. New Directions for Program Evaluation, 4, 75-88.

Rindskopf, D. (1986). New developments in selection modeling for quasi-experimentation. In W. M. K. Trochin (Ed.) Advances in quasi-experimental design and analysis, Number 31, pp. 79-90. San Francisco: Jossey Bass.

Rindskopf, D. (2000). Plausible rival hypotheses in measurement, design, and scientific theory. In L. Bickman (Ed.), Research design: Donald Campbell's legacy, Volume 2 (pp. 1-12). Newbury Park: Sage Publications, Inc.

Roese, N.J. (1997). Counterfactual thinking. Psychological Bulletin, 121, 133-148.

Rosenbaum, P. R. (1984a). From association to causation in observational studies: The role of test of strongly ignorable treatment assignment. Journal of the American Statistical Association, 79 (385), 41-48.

Rosenbaum, P. R. (1984b). The consequences of adjustment for a concomitant variable that has been affected by the treatment. Journal of the Royal Statistical Society, Series A (General), 147 (5), 656-666.

- Rosenbaum, P. R. & Rubin, D. B. (1984). Comment: Estimating the effects caused by treatments. Journal of the American Statistical Association, 79 (385), 26-28.
- Rosenberg, M. (1968). The logic of survey analysis. New York: Basic Books, Inc.
- Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). New York: Academic Press.
- Rubin, D. B. (1974). Estimating Causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66, 688-701.
- Rubin, D. B. (1986). Comment: Which ifs have causal answers. Journal of the American Statistical Association, 81 (396), 961-962.
- Rubin, D. B. (2002, August). Causal inference using potential outcomes. Paper presented at the annual meeting of the American Statistical Association, New York, NY.
- Seavey, C., Katz, P., & Zalk, S. (1975). Baby X: The effect of gender labels on adult responses to infants. Sex Roles: A Journal of Research, 1, 103-109.
- Semin, G.R. & Zwier, S.M. (1997). Social cognition. In J.W. Berry, M.H. Segall, & C. Kagitcibasi (Eds.), Handbook of cross-cultural psychology (Vol. 3) (pp.53-75). Boston: Allyn and Bacon.
- Shadish, W. R. (1995). Philosophy of science and the quantitative-qualitative debates: Thirteen common errors. Evaluation and Program Planning, 18 (1), 63-75.
- Shadish, W. R. (in press). Revisiting field experimentation: Field notes for the future. Psychological Methods.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin.

Sherif, C. W. (1994). Bias in psychology. In A.C. Herrmann & A.J. Stewart (Eds.), Theorizing feminism: Parallel trends in the humanities and social sciences (pp. 117-135). Boulder: Westview Press.

Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M. & Finney, L. J. (1999). Twenty years of PSPB: Trends in content, design, and analysis. Personality and Social Psychology Bulletin, 25 (2), 177-187.

Smith, M. L. (1994). Qualitative plus/versus quantitative: The last word. In C. S. Reichardt & S. F. Rallis (Eds.), The qualitative-quantitative debate: New perspectives, Number 61 (pp. 37-44). San Francisco: Jossey-Bass.

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. Current Directions in Psychological Science, 8, 120-123.

Spirtes, P., Glymour, C. & Scheines, R. (1993). Causation, prediction and search. New York: Springer-Verlag.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52 (6), 613-629.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. Sex Roles: A Journal of Research, 41 (7/8), 509-528.

Sternberg, R. (1988). The Psychologist's companion: A guide to scientific writing for students and researchers, 3rd ed. Cambridge: Cambridge University Press.

Stouffer, S. A. (1972). Willingness for military service. In P.F. Lazarsfeld, A. K. Pasanella & Rosenberg, M. (Eds.), Continuities in the language of social research (pp. 185-189). New York: The Free Press.

Sue, S. (1999). Science, ethnicity and bias: Where have we gone wrong? American Psychologist, 54 (2), 1070-1077.

Tannen, D. (1990). You just don't understand: Women and men in conversation. New York: William Morrow and Company, Inc.

Tetlock, P. E. & Belkin, A. (1996). Counterfactual thought in experiments in world politics. Princeton: Princeton University Press.

Turner, J. & Michael, M. (1996). What do we know about "don't knows"? Or contexts of "ignorance." Social Science Information, 35 (1), 15-37.

Uggen, C. & Piliavin, I. (1998, summer). Asymmetrical causation and criminal desistance. Journal of Criminal Law & Criminology, 88 (4), 1399-1423.

Unger, R. K. & Crawford, M. (1993). Commentary: Sex and gender – the troubled relationship between terms and concepts. Psychological Science, 4 (2), 122-124.

Wainer, H. (1991). Adjusting for differential base rates: Lord's Paradox again. Psychological Bulletin, 109 (1), 147-151.

Walsh, M. E. (1999). Unpacking proxy variables: Cultural factors in adaptation to type II diabetes. Unpublished doctoral dissertation, University of Arizona.

Webb, E.J., D.T. Campbell, R.D. Schwartz, and L. Sechrest. (1966). Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally & Company.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H.T. Reis & C.M. Judd (Eds.), Handbook of research methods in Social and Personality Psychology (pp. 40-84). New York: Cambridge University Press.

Winship, C. & Morgan, S.L. (1999). The estimation of causal effects from observational data. Annual Review of Sociology, 25, 659-706.

Witt, M., Pantula, J, Folsom, R. & Cox, B. (1992). Item nonresponse in 1988. In C.F. Turner, J.T., Lessler & J.C. Gfroerer (Eds.), Survey measurement of drug use: Methodological studies. Washington D.C.: Government Printing Office.

Yee, A.H., Fairchild, H.H., Weizmann, F.& Wyatt, G. (1993). Addressing psychology's problem with race. American Psychologist, 48 (11), 1132-1140.

Zalk, S. R. & Gordon-Kelter, J. (Eds.). (1992). Revolutions in knowledge: Feminism in the social sciences. Boulder: Westview Press.

Zea, M. C., Jarama, S. L. & Bianchi, F. T. (1995). Social support and psychological competence: Explaining the adaptation to college of ethnically diverse students. American Journal of Community Psychology, 23 (4), 509-532.

Zeisel, H. Say it with figures (5th ed.). New York: Harper & Row.

Zeisel, H. & Kaye, D. (1997). Prove it with figures: Empirical methods in law and litigation. New York: Springer.