Theory & Science (2007)

ISSN: 1527-5558

Causation in Quantitative Research Methodologies from Path Modeling, SEM to TETRAD

Chong Ho Yu, Ph.D.
Arizona State University
Email: chonghoyu@gmail.com

Abstract

The objective of this article is to illustrate that from the development of path modeling in the early 20 th century to the recent implementation of the TETRAD approach, the question concerning how causal factors can be identified prior to modeling has been a recurring theme to causal modelers and critics. Wright's path analysis is the predecessor of structural equation modeling (SEM), but Wright was ahead of his time and thus his methodology was sadly ignored by his contemporaries. Nevertheless, during the last two decades Glymour attempted to reinstate causal interpretations for the path model using the TETRAD approach, which is considered an extension to SEM. Interestingly enough, Cartwright's slogan "no causes in, no causes out" (1999, 2000) and Freedman's "missing variable argument" (1997, 1998) in response to Glymour's methodology echo the critics of Wright eighty years ago. Although the philosophical puzzle of causation has not been resolved once and for all, there are substantive improvements in the epistemological aspect of causal modeling, such as the introduction of counterfactual reasoning and manipulation theorem. In addition, abductive-based grounded theory is a promising approach to resolve the issue of variable identification prior to path searching. While Cartwright's dismissal of causal discovery by probabilistic modeling may be exaggerated and even counter-productive, her review of various causal theories is noteworthy, and thus it is the conviction of the author that this type of inter-disciplinary inquiry and synthesis should be further pursued in order to advance epistemology and methodology of causal modeling.

Introduction

The objective of this article is to illustrate that from the development of path modeling in the early 20 th century to the recent revival of the TETRAD approach, the question concerning how causal factors can be identified prior to modeling has been a recurring theme to causal modelers and critics. Wright's path analysis is considered a predecessor of structural equation modeling (SEM), but Wright was ahead of his time and thus his methodology was ignored by his contemporaries. The merits of the path model, which was developed in biology, were appreciated by sociologists in the 1960s, and recently re-introduced back to biology by plant ecologist and biometrician Bill Shipley (2000). However, the path model developed by Wright is considered a "complete" system in the sense that all relevant causal variables have been included. In other words, Wright's so-called causal modeling serves to verify the causal factors and to measure their strength, not to discover causes. Not surprisingly, during the controversy of path modeling in the 1920s, Wright faced rejection from many prominent scholars, including R. A. Fisher. It was argued that the path model assumed a causal structure at the beginning, but without a mechanism for identifying the relevant causal factors, path analysis cannot be considered a true causal model. During the last two decades Glymour attempted to reinstate causal interpretations for the path model using the TETRAD approach. Eighty years after the introduction of path modeling, Freedman (1997, 1998) articulated the "missing variable argument", while Cartwright (1999, 2000) raised the issue of “no cause in, no causes out,” further questioning the viability of finding general causal laws by probabilistic modeling. Although the philosophical puzzle of causation has not been resolved, there are substantive improvements in the epistemological aspect of causal modeling, such as the introduction of counterfactual reasoning and manipulation theorem. While Cartwright’s dismissal of causal discovery by probabilistic modeling may be exaggerated and even counter-productive, her review of various causal theories is noteworthy.

It is my conviction that researchers are not necessarily bogged down by the potential problem of missing relevant variables. Grounded theory based upon abductive reasoning, another technique for discovering causal constructs, is a qualitative method that does not invoke any computation, yet its power for discovering relevant variables is no less than methods using rigorous algorithms. Please note that grounded theory is introduced in an inclusive fashion and it is by no means considered a replacement for quantitative methods.

Path Analysis

In the early 20 th century, biologist Sewall Wright developed a methodology named *path analysis*, regarded as the precursor to structural equation modeling (SEM), in an attempt to enable scientists to assert causal inferences. Simply put, a Pearson’s correlation coefficient is a non-directional or symmetrical covariation. The statement “A and B are significantly correlated” can be pictorially illustrated as “A ↔ B”. In other words, neither “A causes B” nor “B causes A” can be deduced from a correlation coefficient. However, Wright took advantage of the fact that some variables in biology are related in irreversible sequences, and thus one can make deductions of one-way causation. Thus, a path coefficient in the path model has a different notation: A --> B, which implies that there is a causal path from A to B.

Mathematically speaking, a path coefficient was just a standardized regression coefficient, But what distinguishes it from a regression coefficient is that the path coefficient emphasizes causal interpretation rather than purely statistical description. In 1918 Wright used partial correlations to show that while general body size is tied most to variation in bone size, there are unique factors for skull and leg bone dimensions, respectively. It is noteworthy that even in this early work what are “causes” and what are “effects” are not solely identified from the data, but from the prior information brought to the analysis. In 1921, Wright further developed his methodology by employing a set of simultaneous equations. Some critics commented that his claim of causal relationships was ungrounded. In response to his critics, Wright argued that although a coefficient of correlation gives an absolute measure of association in a body of data, a path coefficient can be interpreted only from a particular viewpoint depending on the specification of the model. Wright admitted that his method was identical with those of multiple regression and factor analysis in a closed system, but the path analysis was designed for use in irregular systems with intermingled known and hypothetical variables or known and unknown path coefficients. where conventional methods cannot be applied well.

Path analysis was not well received because some of his contemporaries could not see a way to verify the causal relations that Wright imposed on the model. Niles

(1922) was one of many researchers who did not agree with Wright's approach, but his concern was mainly philosophical rather than mathematical. Following the Pearsonian tradition, Niles regarded "causation" as nothing but mere correlation, and thus for him there is no philosophical ground for giving causation a broader meaning than partial or absolute association. From his perspective, there was no proof that causation is inherent to the laws of nature (Shipley, 2000). In addition, Niles disagreed with Wright on the assumption that a causal system could be initially set up and justified a priori (Denis & Legerski, 2006). Others were opposed to path analysis for methodological reasons. Tukey, the father of the exploratory data analysis, believed that path analysis was good, but not good enough. Comparing path analysis and regression analysis, Turner and Stevens (1959) preferred the latter. Shipley (2000) suggested that Wright's method was rejected for two reasons. First, his orientation was contrary to the philosophical and methodological paradigms of the two prominent schools of thought of the time, namely, Pearsonian and Fisherian. In the Pearsonian school, which emphasized description and categorization, the search for "causation" was simply out of the question. For Fisherians, the central theme of research was to partition true variance and error variance through experimentation; hence, a path coefficient was nothing more than an extension of Pearson's correlation coefficient. Second, Wright's methodology was inadequate in comparison with Fisher's fully developed randomized experimentation. The major shortcoming of Wright's method was that causal relations were assumed into the model, but he did not propose a convincing way to test the hypothesized causal structure.

Structural Equation Modeling

Path analysis had been buried in the history of science for almost three decades. Fortunately, in the 1960s its potential merits were re-discovered by sociologist Hubert Blalock (1964), who at the time was facing problems with analyzing causal relations in sociological studies using partial correlations. Following up on Blalock's work, Duncan recalled a lecture by William Ogburn that had mentioned Burks's (1928) research, which was based upon Wright's work. After reading those archives, Duncan began to apply path analysis to the study of socioeconomic achievement and subsequently to a wider array of sociological topics. In 1970 Duncan and his colleagues organized a conference in Madison, Wisconsin. At the Madison conference Joreskog introduced the idea of Linear Structural Equation, which later was materialized in LISREL, the first software application for SEM.

In 1979 Kenny shocked the methodology community with his book entitled *Correlation and Causality*, a bold attempt to extract causal conclusions from SEM and other multivariate procedures.[1] The book starts with a provocative claim:

> Given the old saying that "correlation does not imply causation," one might wonder whether the stated project of this book—correlational inference—is at all possible. Correlational inference is indeed possible through the application of standard multivariate statistical methods to a stated structural model (p.1).

This theme recurs throughout the book. Like Wright, Kenny did not specify any causal discovery techniques; rather, he argued that before testing the model a causal structure must be pre-specified with reference to prior theories:

> Structural equation models require a blend of mathematics and theory. Although there are many interesting issues in the mathematics of models, the most difficult questions are those that translate theory into equations. This process of translation is called specification. Theory specifies the form of equations (p.22).

In spite of Kenny's good intention of transforming correlation into causation, this ambitious project attracted heavy criticism. For example, Ling (1982) made a strong protest to Kenny's book, arguing that his approach is "a form of statistical fantasy" based upon "faulty fundamental logic" (p.490). Ling asserts that using the path analysis, as illustrated by Kenny, one can never disconfirm a false causal assumption, hence this methodology is "neither science nor statistics" (p.490).

Aware of the theoretical shortcomings of Kenny's approach, James, Mulaik, and Bret (1982) rigorously defended use of SEM as a causal modeling technique by introducing the philosophy of Simon (1952). According to Simon, mathematical and logical equations alone do not imply causal relations, but many researchers often given causal account of the behavior of a system expressed as a set of mathematical and logical relations. Instead, the theory of causal ordering should be employed to infer the causal dependency relations among variables. Some logicians treated connections between variables as nothing more than functional relations. For instance, $Y = a + bX$ can be rewritten as $X = (Y – a)/b$. However, based upon Simon's view of causation, not only do James et al. refuse to view causal modeling expressed by equations as symmetrical functional relations, but also they deny that causation can be defined in terms of logical implications. It is undeniable that the statement "If A then B" logically implies the statement "If not B, then not A." But, if causal relations follow the form of logical implications, this would mean that "A causes B" implies "not B causes not A," which is absurd. Consider this example: The statement "the rain causes Jones to wear his raincoat" does not imply that "Jones not wearing his raincoat causes it not to rain," which is clearly false. Hence, James et al. assert that representation of causation takes the form of an *asymmetrical* functional relation in a self-contained or closed system. In their view, SEM is a type of causal modeling that consists of asymmetrical multiple equations. Put simply, the equation $Y = a + bX$ as a representation of a cause-effect relationship cannot be re-expressed as $X = (Y – a)/b$. The rationale is that theory can play a major role in the formulation of the causal order and can also be used to identify relevant variables in a self-contained system.

From path analysis to SEM, causal modelers rely on theory to formulate a self-contained system, which must initially include all relevant variables. From this perspective, SEM seems to be a tool best utilized to test (confirm/disconfirm) a theorized causal structure rather than to discover one. A structural model is said to be confirmed if the model is consistent with the data. However, data-model fit, in Van Fraassen's term, solely fulfills the criterion of *empirical adequacy*. A theory is considered empirically adequate if and only if everything that it says about *observable* entities is well-established. Van Fraassen (1980) asserts, "Theory draws a picture of the world. But science itself designates certain areas in this picture as observable. The scientist, in accepting the theory, is asserting the picture to be accurate in those areas" (p.57). In actuality, this is an elaborate way to say that the purpose of scientific theory is to "save the appearance," which means apparently the theory can explain the data. However, counting on fitness as the sole criterion to confirm a model, in the eyes of realists, is far from reaching a genuine causal conclusion, in which a causal mechanism underlying the internal structure of the world should be identified. In an attempt to fill-in this philosophical vacuum, Glymour (1982) developed a causal-oriented methodology based on the foundation of SEM, which will be discussed next.

TETRAD

During the past four decades SEM has gained the hegemonic status in multivariate statistics, however, its limitations with causal inferences is still a puzzle. For example, Bentler explains (2005) why SEM can yield a causal conclusion:

> A regression equation in the context of a causal model is called a structural equation, and the parameters, structural parameters. Structural parameters presumably represent relatively invariant parameters of a causal process, and are considered to have more theoretical meaning than ordinary predictive regression weights,

> especially when the regression equation is embedded in a series of simultaneous equations designed to implement a substantive theory. The variables used in the equations must, of course, adequately represent critical substantive concepts, and the model design must be appropriate to the theoretical specification and should include relevant causal variables if at all possible (EQS online manual, Chapter 2).

Careful readers can see that Bentler says nothing more than what Wright had said: Based on prior knowledge of a substantive theory the model can be constructed to represent critical causal factors. But, the initial question about how the model can identify causal factors remains unanswered. To address the issue of causal discovery, researchers in the Philosophy Department at Carnegie Mellon University (CMU) have produced a software module named TETRAD for discovering the causal influences among constructs/variables using automated search algorithms (Glymour,1982).

The search algorithm in TETRAD, as its name implies, utilizes Spearman's tetrad difference equations vanishing tetrads (Hart & Spearman, 1913). Thus, in order for the program to find a subset of measured variables for a factor model, at least four indictors (measured variables) per factor are required. Tetrad refers to the difference between the product of a pair of covariances and the product of another pair among four random variables. For example, if there are four variables, namely, X 1, X 2, X 3, and X 4, there will be three tetrad difference equations:

D1 = s 12 s 34 - s 13 s 24

D2 = s 13 s 24 - s 14 s 32

D3 = s 14 s 42 - s 12 s 43

If the tetrads equal zero, they are called vanishing tetrads, which indicate that the four variables share a common latent factor. [2] In other words, the researcher should obtain zero partial correlations when the model is *linear*. In TETRAD, significance tests are conducted on partial correlations to determine whether two variables are independent given fixed values for some set of other variables. This requirement is called conditional independence, which will be discussed in a later section.

Although the tetrad difference equation was the first attempt to detect latent constructs, it was eventually overshadowed by other techniques such as principal components (Hotelling, 1933), maximum likelihood (Lawley & Maxwell, 1971) and weighted least squares (Browne, 1984). Nonetheless, after the vanishing tetrad approach was revived by Glymour and his colleagues in recent years, many researchers also endorsed it in various applications. For example, when Mulaik and Millsap (2000) defended use of four indicators per factor in their four-step approach for testing a SEM, they praised the tetrad approach for its merits of over-determining the latent variable. To be specific, one can always find a perfect fit between a uni-dimensional factor model with three positively correlated indicators. In this case no test of the single-factor model is possible with this set up. However, four positively correlated variables may not have a single common factor, and as a result, this over-identified common-factor model is testable or refutable.

Counterfactuals

TETRAD's search algorithm, which explores many possible alternate models, can be viewed as a form of manipulation by counterfactual reasoning (Meek and Glymour, 1994). In other words, the prediction of the effect of an intervention on a system is a counterfactual prediction, meaning that it is not a prediction about the existing population, but about a population that does not exist and might never exist (Scheines et al., 1998).

Counterfactual reasoning, as the name implies, is about asking “what-if” questions. When X occurs and Y follows, researchers cannot simply jump to the conclusion that X causes Y. The relationship between X and Y could be “because of,” “in spite of,” or ‘regardless of.” A responsible researcher would ask, “What would have happened to Y if X were not present?” In other words, the researcher does not base one’s judgment solely on the existing outcome, but also other potential outcomes.

The TETRAD methodology is tied to the Fisherian tradition, which endorses use of randomized experiments that often have a quasi-counterfactual aspect. To be specific, the control group gives the information about how Y behaves when X is absent, while the treatment group tells the experimenter about how Y reacts when X is present. Strictly speaking, this comparison is not counterfactual because both the results of the control and treatment groups are actually observed. Rather, the comparison between them is used to make counterfactual inferences in other situations outside the experimental result. Take clinical research as an example. Assume that a drug is given to the treatment group and subjects in the control group receive a placebo treatment. If the health condition of the treatment group patients improves while that of the control group patients deteriorates, potentially resulting in death without proper treatment, it would clearly be unethical for the researcher to carry on the experiment. Rather, the study would halt immediately and counterfactual statements based on the observed result so far, such as the following statement, would be made: “If patients who suffer from disease X do not take Drug A, they would die within three months.”

The counterfactual approach taken by experimenters is limited in two senses. First, according to researchers who strongly embrace the experimental tradition, causal inferences cannot be made from non-experimental data. The preceding example depicts a limited experiment, but in some cases experiments are impossible or entirely unavailable. For example, we cannot use human subjects to experimentally test the impact of smoking on human health. Second, the experimenter can manipulate only few scenarios, but by manipulating many possible models, the researcher may be able to draw causal interpretations from non-experimental data.

Moreover, in the classical Fisherian probability theory, the researcher rejects the null hypothesis because the observed results would be highly improbable compared to other possible outcomes (Howie, 2002). This inferential reasoning based upon comparison across different possible outcomes is clearly counterfactual. Fisher’s randomization exact test also utilizes the same sort of counterfactual logic. The researcher asks what other potential outcomes would result from all other possible scenarios, and the judgment is based on the comparison between the actual outcome and many simulated outcomes (Yu, 2003, 2004, 2006).

Manipulation

In an approach built on Fisher’s legacy, Meek and Glymour (1994) proposed the “manipulation theorem:” Given an external intervention on a variable (A) in a causal model, the researcher can derive the posterior probability distribution over the entire model by simply using the conditional probability distribution of A. If this intervention is strong enough to set A to a specific value, the researcher can view the intervention as the only cause of A. Nothing else in the model needs to be modified, as the causal structure of the system remains unchanged. In other words, if counterfactual scenarios can be generated under such assumptions by manipulating values and variables in equations, causal inferences can be made.

Although Woodward and Glymour are not in the same research camp (and indeed Glymour [2004] criticizes Woodward on some points), they share much common ground concerning manipulation and intervention. Experimentalists who follow the Cook and Campbell (1979) approach regard so-called "causal inferences" in non–experimental settings as illegitimate because manipulating equations in modeling should not be treated as real intervention. However, manipulating equations is like conducting thought experiments in a theoretical sense (Woodward, 2003). Conventional wisdom suggests that one may infer causal relationships in practical, experimental, and "applied" science contexts, in which physical changes of variables can be manipulated and observed. However, it seems counter-intuitive to think of causal relationships in "pure" or theoretical sciences in this way. Thought experiments have been widely used in theoretical physics, game theory, economics, and theoretical evolutionary biology (Cooper, 2005). It is problematic to maintain two distinct notions of causation, one in practical contexts and the other in theoretical contexts. It is possible that due to technological advances, some studies in the past that were confined to thought experiments can eventually be scrutinized in physical experiments. Take the Einstein-Podolsky-Rosen (EPR) experiment as an example. Although originally it was conceptualized as a thought experiment by Einstein, Podolsky, and Rosen to argue against "entanglement" in quantum mechanics, later technological advancement enabled scientists to carry out actual experiments regarding entanglement and locality. It was demonstrated that non-local effects on particles were present and the EPR's claim of "locality" of influences was rejected (Aczel, 2003).

The meaning or role of causal claims should be the same in theoretical and practical situations. Further, a process or event could still qualify as an intervention even if it does not involve human action (Woodward, 2000, 2001, 2003). In other words, a purely "natural" process involving no animate beings at all can qualify as an intervention if causal information is embedded. This type of research is often described by scientists as a "natural experiment."[3] Moreover, even when manipulations are carried out by human beings, it is the causal features of those manipulations that matter for recognizing and characterizing causal relationships. In experiments, human intervention occurs in the real world. In the mathematical world, intervention or manipulation happens in a counterfactual fashion, or in "other possible worlds." The intervention yields answers to questions like "what would happen to Y if X1 were added to the model and the coefficient of X2 were down–weighted?" In this case, whether or not the interventions that set the value of Xs and Y are carried out by human beings and whether or not they have in fact taken place is irrelevant (Hausman & Woodward, 1999).

Intervention as it occurs using TETRAD fits this idea. Human intervention in experiments does not create causal information or make the data ready for causal structure. Causal properties have already been embedded in the subject matter and experimental control is a way to reveal the causal information. If the data are non–experimental, causal characteristics are still within the data model. Mathematical intervention, by the same token, makes the causal relationships more obvious, if there are any. In TETRAD, causal structure is represented in a system of equations. When the researcher changes the variables and/or the coefficients of the equations, he/she is changing the mechanism(s) or relationship (s) represented by it. We can view this as a matter of intervening on the dependent variable in the equation so that the value of that variable is now fixed by the intervention rather than by the variables that previously determined its value (Woodward, 1999).

During the 1920s, Wright's path modeling was under severe criticisms for the model is treated as a close system based on the assumption that all relevant variables have been included. Interestingly enough, criticisms made by contemporary mathematician Freedman and philosopher of science Cartwright against causal modeling echo the controversy of path modeling eighty years ago.

Missing variable argument

Can TETRAD modeling yield causal conclusions in the absence of inputted causal information? In Freedman's view, if some variables that play crucial causal roles are omitted at the beginning, the selection algorithms may not be capable of recovering the omitted variables and the researcher may get the wrong conclusion, misled by the so-called best fit. Usually, the impact of random variation and omitted variables on the model are represented by the error terms. The errors are assumed to be drawn independently from a Gaussian distribution. Generally, the error distribution is not empirically identifiable outside the model, so it cannot be studied directly without the model. The error distribution is an imaginary population and the errors are treated as if they were a random sample from this imaginary population. Although structural equation models seem sophisticated, the same old problems have been swept under the carpet, because random variation due to unaccounted variables is represented in the same old way (Berk & Freedman, 2003). Simply put, when relevant variables are omitted, no one can tell how much error is in the model.

Freedman's argument is essentially "If you don't know enough about causes, you cannot conduct causal modeling." But what is considered "enough"? Actually, it is totally acceptable to miss some variables and then expand the system by adding more variables later. As a matter of fact, in the history of science even successful theories could not include all relevant variables and accurately predict every aspect of a phenomenon. Consider the example of cognitive assessment models. In recent years psychometricians have recommended including cognitive psychology in assessment (National Research Council, 2004). Rather than hastily inventing test items, item authors are advised to conduct a careful task analysis by decomposing a complex task into sub-tasks in a Markovian process: the success of completing the step, A1, depends on the pre-requisite, A2; A2 depends A3 and A4; A3 depends on A5 while A4 depends on A6 and A7; and so on. These subtasks should be mapped to a series of test questions that specifically reflect the required mental constructs for accomplishing the subtasks, so that the exam can perform diagnosis based on conditional probabilities of the failure or success of each subtask. In other words, item authors must understand the cognitive process of the students well enough to create a cognition-item map. However, it is extremely difficult, if not impossible, for item authors, psychometricians, cognitive psychologists, and educational psychologists to include all relevant variables in such complicated cognitive processes. There will always be missing variables, but the expectation is that as this new assessment movement grows, more and more variables will be added into the cognitive assessment models. Consider assessment again. A Rasch model or a one-parameter item response theory model completely ignores the discrimination and guessing parameters. The discrimination parameter acknowledges the fact that the same item might have different difficulty levels to examinees of different ability, while the guessing parameter takes into account the fact that there is a chance that examinees who know nothing about the subject matter could score the item by guessing. One may argue that these are important variables, yet psychometricians are still able to conduct meaningful assessment with a Rasch model or a one-parameter item response theory model.

"No causes in, no causes out"

Another skeptical view is expressed by a slogan introduced by Cartwright (1999): "no causes in, no causes out" (p.39). In her view, there is no way to get causal information from equations and associations. New causal knowledge must be built from old, empirical causal knowledge. In other words, the empiricist's rule embraced by Cartwright is that the relevant data are the data that will inform us about the truth or falsity of the hypothesis, given the other known facts. Glymour et al. include all possible combinations of variables and paths in the model and then the irrelevant ones are eliminated. Cartwright argued that if relevant variables and genuine causes are not included at the beginning, then this elimination approach is useless. For these reasons, Cartwright strongly criticized Glymour et al.'s theory:

> Because Glymour, Scheines, Kelly, and Spirtes employ the hypothetico–deductive method, they must proceed in the opposite order. Their basic strategy for judging among models is two–staged: first list all the relevant relations that hold in data, then scan the structures to see which accounts for the greatest number of these relations in the simplest way. That means that they need to find some specific set of relations that will be relevant for every model. But, from the empiricist point of view, no such thing exists (p.78).

In questioning the applicability of Causal Markov Condition (CMC), a crucial assumption of TETRAD, Cartwright (1999) used a classical example to argue that researchers may take the risk of confusing a co–symptom with a cause: R. A. Fisher's hypothesis that smoking does not cause lung cancer. Rather, smoking and lung cancer are caused by a common cause, namely a special gene that increases the tendency to smoke and to get cancer. Not surprisingly, Cartwright asserted that to investigate a hypothesis like this, one must conduct a randomized experiment instead of counting on CMC and mathematical intervention of non–experimental data.

The statement "no causes in, no causes out" is tied to Cartwright's notion of *nature's capacities* in a *dappled world* (1999). The so-called dappled world is a world that is not governed by universal laws. Rather, every phenomenon is a consequence of many interacting parts, thus talking about universal causal laws doesn't make sense in a dappled world. As a remedy, she proposes thinking of nature's capacities in different situations. Cartwright argues that statistical methods and probability can support causal inferences if and only if the probability of the effect given the presence of the alleged cause is higher than the probability of the effect given the absence of the cause in all conditions. But Cartwright cites Simpson's Paradox (Simpson, 1951), in which a conclusion drawn from aggregate data is contradicted by the conclusion drawn from the contingency table based upon the same data, to deny that this condition can be met. For instance, in England once a 20–year follow–up study was conducted to examine the survival rate and death rate of smokers and non–smokers. The result implied a significant positive effect of smoking because only 24% of smokers died compared to 31% of non–smokers. However, when the data were broken down by age group in a contingency table, it was found that there were more older people in the non–smoker group (Appleton & French, 1996). Based on Simpson's Paradox, Dupre and Cartwright (1988) suggested that there are only probabilistic capacities, but no probabilistic causal generalizations at all. Hence, according to Cartwright, we have to know what all other interacting causes are and how they work in order to conduct inquiry about statistical and probabilistic causation in specific situations. Clearly, Cartwright and Glymour's ideas are quite diametrically opposed.

To be more elaborate, "no causes in, no causes out" means "no all inter-entangling causes in, no causal structure out." Knowledge of interacting causes comes from sources other than statistical data, and once they are known, statistics becomes irrelevant because the causes are obvious. Those interacting causes that render universal causal laws useless are usually the missing variables that have not been included in the model. Along this line, Cartwright's argument is similar to Freedman's.

Cartwright (1999) cites an example introduced by one of the founders of the Vienna Circle, Otto Neurath, to illustrate why prediction and causal explanation appealing to universal laws are doomed to fail. She explains that if you drop a ten thousand dollar bill in St. Stephen's square on a windy day and expect the bill would behave according to the second Newtonian law of mechanics, you will fail to predict the trajectory. Indeed, the bill will be swept away by the wind and no physicist can predict where it will land.

While the above point sounds philosophically interesting, in *practice* physicists and engineers do not subscribe to Cartwright's idea. In the late 1950s and early 1960s, Americans were deeply concerned with the "missile gap" and space race after the USSR successfully tested their Inter-Continental Ballistic Missile (ICBM) system and launched the first satellite, Sputnik, by their powerful rocket, R7. When President Eisenhower urged US scientists to catch up in research on long range missiles, would he have felt better if an American scientist said, "Don't worry, the trajectory of missile path could not be explained by universal laws. Many other causal factors that are unknown to us might deflect the route and thus no one could predict exactly where the Russian missile would hit?" Unlike the ten thousand dollar bill, a missile would never be affected by the strong wind. When President Kennedy set his goal of landing a man on the moon by the end of 1960s, would he have been discouraged if a scientist said, "It is impossible to do space travel. There are so many unknown factors beyond the terrestrial domain that no physicist could be certain that our spacecraft can land on the planned location"? To paraphrase JFK's speech: Researchers look for causal modeling not because it is easy, but because it is hard!

Many physicists and philosophers of science question whether the example of the un-predictable ten thousand dollar bill can negate the universal applicability of Newtonian mechanics and other laws in physics. Actually, the bill's deviation from a free-fall trajectory can be explained by other forces, such as the wind and air resistance (Hoefer, 2003). It is important to distinguish a *phenomenologically dappled world* from an *ontologically dappled world*. In the former, even if fundamental laws exist, the world appears to be disunified due to the limitations of human knowledge. In the latter, the world is fundamentally disunified, no matter how it appears to us (Hohwy, 2003). If there are laws other than the Newtonian mechanics that can explain the movement of a bill on a windy day, then this example should be put under the first category. More specifically, a phenomenologically dappled world does not necessarily imply an ontologically dappled one.

At the fundamental level, laws tend to have few exceptions. For example, a hydrogen atom in a spectrometer is much the same as a hydrogen atom floating in your living room. The existence of a stable state, in which the proton and electron are bound to each other spatially, yet never collapse, is another good example (Hoefer, 2003). Aside from the example of the flying bill, Cartwright cites many other examples from physics, however, all of the theories that she lists under classical mechanics for demonstrating the failure of fundamental laws are indeed sub-theories of other more encompassing theories (Smith, 2001).

The dominant view regarding the stability of fundamental elements and universal laws among physicists is that the concepts of electrons, protons, neutrons, neutrinos, and quarks, as well as the theories of quantum mechanics, general relativity, and special relativity are *here to stay* (Hacking, 1999). Whatever revolutionary changes physics may introduce in the future, it is very unlikely that the basic ideas encompassing electrons, protons, neutrons, neutrinos, and quarks will be found false and their causal relations rejected.

In fairness to Cartwright, her notion of nature's capacities in a dappled world is not just to argue against general causal laws; rather, her motivation is to counteract the trend of seeking for complete and universal explanations of every physical phenomenon. String theory is a typical example of this bold attempt (e.g. Greene, 1999, 2004). According to Cartwright (2000), the vision of completeness is typically combined with some kind of reductionism. In the past, reductionism was vertical or downward, in which laws in one discipline were said to be explained by those in another discipline at a more fundamental level. For example, biology is said to be reduced to chemistry whereas chemistry can be reduced to physics. Today, reductionism is horizontal or cross-wide, in which diverse phenomena studied by different disciplines are governed by a few fundamental laws (Ruphy, 2002). Interestingly enough, while Cartwright is critical of any form of reductionism, her idea is a form of upward expansionism. She uses physics as a starting point to illustrate the non-existence of universal laws, and then gradually expands this notion beyond physics to social sciences and other disciplines.

Further, while it is reasonable to question the completeness thesis, it does not follow that "without complete knowledge, no causal modeling is possible." It is understandable to warn causal modelers about the danger of over-generalization, especially in policy studies, [4] but it is doubtful whether objecting that a model may leave out some genuine causes or relevant variables and consequently rejecting the method could help scientific progress at all. First, who could affirm that all relevant variables are included in the model except an omnipotent God? Second, is it really necessarily to include all relevant variables? In defense of his standpoint, Glymour (1999) wrote,

> Cartwright is perhaps correct that the whole truth about anything is very complex; but, quite properly, science is seldom interested in the whole truth, or aided by insistence upon it. In my view, an inquiry that correctly found the causes of most of the variations in a social phenomenon and neglected small causes would be a triumph" (p.59).

Glymour clearly appreciates the difficulty in uncovering truth in any situation, however, feels that useful knowledge can still be gained from such "imperfect" research.

Discussion

As illustrated above, the issue of identifying causal variables has been repeated by critics, such as Niles, Fisher, Ling, Freedman, and Cartwright, throughout the years since the development of path modeling, SEM, to TETRAD. Denis and Legerski (2006) are correct that the original path model introduced by Wright and revived by Blalock assumes prior causal factors and thus it is nothing more than a fancy regression model. However, Glymour (2001) also treat regression-based approaches as the wrong way for causal inferences. In the CMU group, path modeling has evolved to a more sophisticated methodology based upon conditional independence rather than regression coefficients. More importantly, the theories of counterfactual reasoning and manipulation seem promising. It is true that the same question has been repeated, but the answers are getting better and better.

One of the reviewers of this manuscript doubted whether there is, or ever will be, a quantitative solution to the problem of causality in the social sciences. In my view, a purely quantitative approach is doomed to be trapped by the same critiques of path modeling and Freedman and Cartwright's criticism of TETRAD: "How can the researcher come up with a list of relevant variables in the first place?" Rather, a mixed-method, in which an initial qualitatively-oriented abductive reasoning for theory generation and a subsequent quantitative-based path searching are integrated, should be considered by causal modelers. In the context of causal discovery, abduction (Peirce, 1934/1960) is indispensable at the stage of identifying and categorizing variables and factors, which are the building blocks of causal modeling. Without this essential component, neither automated data mining nor conventional hypothesis testing could build a meaningful causal model.

The function of abduction is to look for a pattern in a surprising phenomenon and to suggest a plausible hypothesis. The following example illustrates such function:

- The surprising phenomenon, B, is observed.
- But if A were true, B would be a plausible explanation.

- Hence, there is a reason to suspect that A might be the explanation.

Using deductive reasoning, the preceding example is akin to the fallacy of affirming the consequent. Consider this example. It is logical to assert that "It rains; if it rains, the ground will be wet; hence, the ground is wet." But any reasonable person can see the problem in making statements like: "The ground is wet; if it rains, the ground will be wet; hence, it rained." Nevertheless, in Peirce's logical framework this abductive form of argument is entirely acceptable, especially when the research goal is to discover plausible explanations for further inquiry (de Regt, 1994). In order to yield a set of plausible explanations, abduction is usually formulated in the following mode:

- The surprising phenomenon, X, is observed.
- Among hypotheses A, B, and C, A is capable of explaining X.
- Hence, there is a reason to pursue A.

Although abductive reasoning seems problematic from a deductive standpoint, it is a legitimate scientific methodology, which appears in what's known as "reverse engineering." Inquiry in evolutionary biology utilizes reverse engineering, or abduction, because biologists trace back the causal history of evolution given the consequences (existing species and fossil records) (Kleiner, 2003).

The preceding discussion outlines the principle of abductive reasoning. Specific implementations of the abductive principle can be embodied in certain qualitative methodologies, such as grounded theory (GT). Indeed, grounded theory, as a non-quantitative method, is a viable way to generate theories that explain the qualitative data patterns from which they are derived (Haig, 2005, 1995). Take the construct "intelligence" as an example. Early psychometricans made a conjecture that there was a single G factor that subsumed all aspects of human intelligence; all standardized tests of academic aptitude or achievement measure this general factor to some degree, but IQ tests expressly measure it more accurately. To counteract this notion, Gardner (1993, 1996) developed the multiple-intelligences model, which proposes that human intelligences have at least eight dimensions, namely, linguistic, logical-mathematical, interpersonal, intrapersonal, artistic-spatial, musical, kinesthetic, and naturalist. Contrary to popular belief, this model was not developed using factor analysis. Instead, Gardner used a "subjective" and qualitative approach to identify those latent constructs.

GT is one of the qualitative approaches commonly used for discovering new concepts and proposing new theories (Glaser & Strauss, 1967; Dey, 1999). It was named "grounded theory" because its core idea is to ground the theory on the data rather than taking a pre-established theoretical framework for granted. Interestingly enough, the inferential process of grounded theory is like *reverse engineering*, as discussed above. In GT, the relevant constructs or variables are extracted from the phenomenon under study. GT emphasizes exploratory work in the sense that concept identification and theory generation should be an iterative process; this process is stopped if and only if the category or the concept is saturated, which means collecting new data can no longer add anything new to the existing category. What is involved in this iterative process is constant comparison. At the beginning, the researcher compares interview (or other data) to interview (or other data) in order to extract a recurring theme for theory development. After a theory has been developed, the researcher compares data to theory. Unlike factor analysis where a structured questionnaire or exam is always used, the interview conducted by the grounded theorists is usually unstructured. While a factor analyst has some preconceived ideas about what is expected to be observed, a grounded theorist, on the contrary, avoids developing a list of preconceived codes before the project starts. In addition, unlike quantitative research, there is no rigid sampling scheme in GT. The researcher can choose to expand or narrow down the sample pool based upon the emerging theory. Because the sampling scheme is driven by the emerging theory, it is termed “theoretical sampling.” Further, in quantitative research outliers or misfits are always excluded from the analysis, and very few researchers ask why some observations cannot be fit into the overall pattern. On the contrary, “strange” data are never considered a source of embarrassment by grounded theorists, but an excellent occasion for what they may contribute to expansion, refinement, and enrichment of the emerging constructs (Glaser, 1978).

At first glance, this approach of concept identification is very loose in procedure and highly speculative. Many qualitative researchers, including grounded theorists, subscribe to the worldview of constructivism, in which reality is viewed as a social construction based upon human perspectives. Cutcliffe (2000) argued that constructivism is not a license for the grounded theorist to freely invent concepts and categories. Rather, what it does is to legitimize the researcher's creativity as an integral part of the grounded theory inductive process; liberating the restrictions on the researcher's tacit knowledge that discounting such knowledge creates. While Cutcliffe is correct to emphasize the creative aspect of GT, it is problematic to view GT as an inductive process only, because new ideas should arise from abduction as well as induction. Roughly speaking, induction is a mode of reasoning by seeing more of the same “kind,” but there is no pre-conceived “kind” at the initial stage of GT. Usually, induction happens at the later stage of the inquiry. For this reason, Haig (1995), Rennie (1999, 2000) and I all regard GT as an abductive method. In induction, the researcher comes to a conclusion that a stable construct is identified if the same pattern recurs many times, but in order to recognize whether observed phenomena converge into a pattern, the research question must be stated in advance. However, grounded theorists insist that we do not have to prepare an articulated problem in advance of inquiry, rather researchers may come to their problems at any point in the research process and new concepts can be proposed as new data update our understanding of the phenomena. This open-ended qualitative approach may be a viable way to face the century-old challenge against causal modeling: “How can the researcher come up with a list of relevant variables in the first place?”

Last, causal modeling is no longer a mere methodological issue that can be investigated by statisticians alone. As illustrated above, some insightful theories for causal modeling and their counter-arguments are developed by philosophers, such as Simon, Glymour, Woodward, and Cartwright. Qualitative researchers can also make significant contributions to causal discovery in terms of variable identification and theory generation. In an attempt to examine whether causal theories in various fields share a common thread, Cartwright (2006) compared Suppes’s probabilistic theory of causality, Bayes-net theories, Granger causality, modularity accounts, manipulation accounts, invariance accounts, natural experiments, causal process theories, the efficacy account, and counterfactual accounts. However, owing to her insistence that in a dappled world causality in terms of probability is not viable, her theory of causal theories does not seem to provide a usable guideline to philosophers and social scientists. Hence, further endeavors in studying various causal theories are needed.

In physics, the superstring theory is under development as a unifying theory of all other theories. But, this type of unification is not what I propose here, nor do I recommend formulating a theory of causal theories to rank diverse causal modeling approaches and to pick the best out of the lot. Different disciplines have different problems in causal modeling. For example, for psychologists the issue is latent constructs whereas for biologists the major concern is fundamental units.[5] A taxonomy that specifies which causal discovery approach is appropriate given which conditions would be desirable. However, it is unlikely for a single scholar to be well-versed in every causal discovery methodology in different fields. Therefore, it will be more fruitful to engage in inter-disciplinary dialogs and collaborations than to try to single-handedly exhaust all causal modeling methods.

Acknowledgements

Special thanks to Dr. Brad Armendt, Dr. Bernard Kobes, Dr. Samuel Green and Mr. Charles Kaprolet for their valuable input to this article.

End Notes

1. Kenny (2004) posted a revised version on his website. The main points that are cited in this chapter remain in the online version. Kenny said, "The revised edition is essentially the same as the original 1979 book. I have corrected the errors that I knew of. I have added a few new references. However, in no way this revised edition should be considered an updated version of the manuscript" (p.vi).

2. Take Spearman's G factor as an example. Zero tetrad differences among measures of reading and math aptitude led to the formulation of the hypothesis that a single common cause, namely, general intelligence, was responsible for performance on all four pyschometric instruments. In other words, if the correlation of a pair of variables is not different from another pair and this pattern holds in all pairwise comparisons, the researcher has no reason to suspect that there is more than one cluster of variables or that there is more than one construct behind those variables. This implication of a common cause holds no matter what values the linear coefficients may have, and no matter what the joint distribution of variables may be. Spearman's tests always assumed that the measured variables are jointly normally distributed. After the causal structure was established, the linear coefficients could be estimated from the data (Glymour, 1999). While checking normality and linearity are not difficult, the null hypothesis is that the tetrad differences are zero in the population, and a certain minimum sample size is required to get an accurate probability estimates based on this hypothesis. But, according to Shipley (2000), no one has formally studied the asymptotic requirement for the vanishing tetrad test.

3. Woodward (2001) cites this example: It has been argued that the stability of planetary orbits depends on the dimensionality of the space-time in which they are situated. Such orbits are stable in a four-dimensional space-time but would be unstable in a five-dimensional space-time. The above claim fits well with the idea that causal explanations provide answers to the 'what-if" question. However, the causal structure of the spatial-temporal dimensions and planetary orbits has been established before the existence of human beings, and even though there are human beings in the universe now, no one could physically manipulate the spatial-temporal dimensions.

4. Cartwright (1999) is very concerned with using causal models developed by Glymour to drive policy-decision. Because every situation is unique, to develop a sound policy we need to know not only what causal relations hold, but what will happen to them when we undertake changes. Cartwright also worries that the quest for fundamental laws is a blind faith that has led to detrimental social consequence. For example, she disagrees with the take-over of genetics as the dominant approach to try to cure diseases like breast cancer. She is afraid that women are dying of breast cancer because treatments other than gene therapy, with good empirical supports, are ignored or under-funded. In brief, Cartwright's theory has a social dimension, not just epistemic.

5. Take evolutionary biology as an example. The concept of "character" is essential to the study of evolution in the context of homology, which is a study of "sameness." While tracing links between species, a biologist must ask a question like "How do I recognize the character in species B that corresponds to the one I know from species A?" This question leads to another question: "What are the

natural units that organisms are composed of?” It would be crazy for a biologist to make a molecule-to-molecule comparison between two species. The biologist must choose some “kinds” or “characters” that are beyond the molecular level (Wagner, 2001). Most latent constructs in social sciences are psychological and cannot be directly observed, while characters and kinds in biology can be observed, but selecting which one as the basic unit of analysis is challenging.

References

Aczel, A. D. (2003). *Entanglement: The unlikely story of how scientists, mathematicians, and philosophers proved Einstein's spookiest theory*. New York: Penguin Group.

Appleton, D. R., & French, J. M. (1996). Ignoring a covariate: An example of Simpson’s paradox. *American Statistician, 50*, 340–341.

Bentler, P. (2005). EQS. [Computer software and manual] Retrieved October 15, 2005, from http://www.mvsoft.com/contactus.htm

Berk, R.A., & Freedman, D. (2003). Statistical assumptions as empirical commitments. In T. G. Blomberg & S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger* (2nd ed., pp. 235–254). New York: Aldine.

Blalock, H. M. (1964*). Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press.

Browne, M. W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Burks, B. S. (1928). The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. *The Twenty-Seventh Yearbook of the National Society for the Study of Education, 27,* 219-316.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cartwright, N. (1999). *The dappled world.* Cambridge: Cambridge University Press.

Cartwright, N. (2000). Against the completability of science. In M. W. Stone (Ed.). *Proper Ambition of Science* (pp. 209-223). London: Routledge.

Cartwright, N. (2006). Where is the theory in our “theories” of causality? *Journal of philosophy, 103,* 55-66.

Cook, T. D., & Campbell, D. T. (1979). *Quasi–experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Cooper, R. (2005). Thought experiment. *Metaphilosophy, 36*, 328-347.

Cutcliffe, J. (2000). Methodological issues in grounded theory. *Journal of Advanced Nursing, 31*, 1476-1484.

de Regt, H. (1994). *Representing the world by scientific theories: The case for scientific realism*. Tilburg, The Netherlands: Tilburg University Press.

Denis, D. & Legerski, J. (2006). Causal modeling and the origins of path analysis.

Theory & Science, 7(1). Retrieved June 16, 2006 from http://theoryandscience.icaap.org/content/vol7.1/denis.html

Dey, I. (1999). *Grounding grounded theory: Guidelines for qualitative inquiry*. San Diego, CA: Academic Press.

Dupres, J., & Cartwright, N. (1988). Probability and causality: Why Hume and indeterminism don't mix. *Nous, 22*, 521–536.

Freedman, D. (1997). From association to causation via regression. *Advances in Applied Mathematics, 18,* 59–110.

Freedman, D. (1998). Rejoiner to Spirites and Scheines, In Vaughn R. McKim & Stephen P. Turner, (Eds.), Causality in crisis? *Statistical methods and the search for causal knowledge in the social sciences* (pp. 177-182), Notre Dame, IN: University of Notre Dame Press.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligence*. New York: Basic Books.

Gardner, H. (1996). *Leading minds: An anatomy of leadership*. New York: Basic Books.

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. New York: Aldine.

Glymour, C. (1982). Casual inference and causal explanation. In R. McLaughlin (Ed.), *What? Where? When? Why? Essays on induction , space, and time, explanation* (pp. 179–191). Boston, MA: D. Reidel Publishing Company.

Glymour, C. (1999). Rabbit hunting. *Synthese, 121,* 55–78.

Glymour, C. (2001). Mind's arrows: Bayes and Graphical causal models in psychology. Cambridge, MA: MIT Press.

Glymour, C. (2004). Book review: Making Things Happen: A Theory of Causal Explanation. *British Journal for the Philosophy of Science, 55*, 779-790.

Greene, B. (1999). *The elegant universe: superstrings, hidden dimensions, and the quest for the ultimate theory*. New York: Vintage.

Greene, B. (2004). *The fabric of the cosmos: Space, time, and the texture of reality*. New York: Vintage Books.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Haig, B. (1995). Grounded theory as scientific method. In A. Neiman (Ed.), *Philosophy of education 1995: Current issues* (pp. 281-290). Urbana: University of Illinois Press.

Haig, B. (2005). An abductive theory of scientific method. *Psychological Methods, 10*, 371-388.

Hart, B., & Spearman , C. (1913). General ability, its existence and nature. *British Journal of Psychology, 5,* 51–84.

Hausman, D., & Woodward, J. (1999). Independence, invariance , and the

Causal Markov Condition . *British Journal of Philosophy of Science, 50,* 521–583.

Hoefer, C. (2003). For fundamentalism. *Philosophy of Science, 70*, 1401–1412.

Hohwy, J. (2003). Capacities, explanation and the possibility of disunity. *International Studies of Philosophy of Science, 17*, 179-190

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 4170441, 498–520.

Howie, D. (2002). Interpreting probability: Controversies and developments in the early twentieth century. Cambridge, UK: Cambridge University Press.

James, L. R., Mulaik, S., & Bret, J. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage Publications.

Kenny, D. (1979). *Correlation and causality*. New York: John Wiley.

Kenny, D. (2004). *Correlation and causality*. Retrieved August 31, 2006, from http://davidakenny.net/cm/cc.htm

Kleiner, S. (2003). Explanatory coherence and empirical adequacy: The problem of abduction, and the justification of evolutionary models. *Biology and Philosophy 18*, 513–527.

Lawley, D. N., & Maxwell, A. E. (1971). Factor analysis as a statistical method. London: Butterworth.

Ling, R. (1982). Review of "Correlation and causation" by David Kenny. *Journal of American Statistical Association, 77*, 481–491.

Meek, C., & Glymour , C. (1994). Conditioning and intervening. *British Journal of Philosophy, 45,* 1001–1021.

Mulaik, S., & Millsap, R. (2000). Doing the four–step right*. Structural Equation Modeling, 7,* 36–73.

National Research Council. (2004). *Knowing what students know: The science and design of educational assessment.* Washington DC: National Academic Press.

Niles, H. E. (1922). Correlation, causation and Wright's theory of "path coefficients." Genetics , 7, 258-273.

Peirce, C. S. (1934/1960). *Collected papers of Charles Sanders Peirce*. Cambridge: Harvard University Press.

Rennie, D. (1999). A matter of hermeneutics and the sociology of knowledge. In M. Kopala & L. A. Suzuki (Eds.), *Using qualitative methods in psychology* (pp. 3–14). Thousand Oaks, CA: Sage Publications.

Rennie, D. (2000). Grounded theory methodology as methodical Hermeneutics: Reconciling realism and relativism. *Theory & Psychology, 10*, 481–502.

Ruphy, S. (2002). Is the world really "dappled"? A Response to Cartwright's charge against "cross-wise reduction." *Philosophy of Science, 70*, 57–67.

Scheines, R., Spirtes, P., Glymour , C., Meek, C., & Richardson, T. (1998). The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research, 33,* 65–117.

Shipley, B. (2000). Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference. Cambridge: Cambridge University Press.

Simon, H. (1952). On the definition of causal relations. *Journal of Philosophy, 49*, 517-528.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Ser. B., 13*, 238–241.

Smith, S. R. (2001). Models and the unity of classical physics: Nancy Cartwright's dappled world. *Philosophy of Science, 68*, 456-475.

Turner, M. E. & Stevens, C. D. (1959). The regression analysis of causal paths. *Biometrics, 15,* 236-258.

van Fraassen, B. (1980). *The scientific image.* Oxford: Clarendon Press.

Wagner, G. (2001). (Ed.). *The character concept in evolutionary biology*. San Diego, CA: Academic Press.

Woodward, J. (1999). Causal interpretation in systems of equations. *Synthese, 121,* 199–247.

Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal of Philosophy of Science, 51,* 197–254.

Woodward, J. (2001). Causation and manipulability. *Stanford Encyclopedia of Philosophy.* Retrieved October 13, 2005, from http://plato.stanford.edu/entries/causation –mani/

Woodward, J. (2003). *Making things happen.* Oxford: Oxford University Press.

Yu, C. H. (2003). Resampling methods: concepts, applications, and justification. *Practical Assessment, Research & Evaluation*, 8(19). Retrieved November 7, 2005, from http://PAREonline.net/getvn.asp?v=8&n=19

Yu, C. H. (2004). Advance in Monte Carlo Simulations and robustness study and their implications for the dispute in philosophy of mathematics. *Minerva, 8,* 62-90. Retrieved September 1, 2006 from URL: http://www.mic.ul.ie/stephen/vol8/montecarlo.pdf

Yu, C. H. (2006). *Philosophical foundations of quantitative research methodology*. Lanham, MD: University Press of America.