Watching Social Science: The Debate About the Effects of Exposure to Televised Violence on Aggressive Behavior

Bruce Glymour, Clark Glymour and Maria Glymour

American Behavioral Scientist 2008; 51; 1231
DOI: 10.1177/0002764207312020

American Behavioral Scientist
Volume 51 Number 8
April 2008 1231-1259

10.1177/0002764207312020
http://abs.sagepub.com
hosted at
http://online.sagepub.com

Watching Social Science

The Debate About the Effects of Exposure to Televised Violence on Aggressive Behavior

Bruce Glymour
Kansas State University, Manhattan
Clark Glymour
Carnegie Mellon University, Pittsburgh, PA
Maria Glymour
Harvard School of Public Health, Boston, MA and
Mailman School of Public Health at Columbia University, NY

The evidence that exposure to media violence causes later aggression derives largely from observational (nonexperimental) studies augmented by short-term experimental studies. The authors review some of the difficulties in causal inference from observational, longitudinal data; examine the extent to which these seem relevant to the empirical work on exposure to televised violence published to date; and present a reanalysis of data from an especially influential study to address one of the more serious limitations of existing analyses. They conclude that the data give evidence that there is likely, although not certainly, a causal connection between exposure to televised violence and adult aggression. The authors close with a brief discussion of policy interventions designed to reduce exposure to violent TV.

Keywords: *causal inference; epidemiologic methods; causal graphical modeling; automated search; televised violence; aggression*

If then we adhere to our original notion and bear in mind that our guardians, setting aside every other business, are to dedicate themselves wholly to the maintenance of freedom in the State . . . they should not depict or be skilful at imitating any kind of illiberality or baseness, lest from imitation they should come to be what they imitate. Did you never observe how imitations, beginning in early youth and continuing far into life, at length grow into habits and become a second nature, affecting body, voice, and mind?
And therefore let us put an end to such tales, lest they engender laxity of morals among the young.

—*Plato (*Republic*, Book III)*

Causal conclusions in sociological studies are of interest largely because they are informative about the range of interventions that would or would not change behavioral variables of interest. If we were to know that watching violent television in childhood causes adult aggression, we might be moved to change social policy, in any of a number of ways, so as to change childhood exposure to television violence. But the social benefits and costs of policy interventions are generally difficult to estimate, and the social costs of policy interventions are often independent of the effectiveness of those interventions. When those costs are reasonably expected to be large, ineffective policy interventions are immoral. On the other hand, failing to report the most plausible causal alternatives on matters of substantial personal and social relevance may be similarly irresponsible. Parents and policy makers must make decisions that will influence children's exposure to TV viewing, and it is the proper role of the research community to provide the best possible information, along with an honest assessment of uncertainty, on which to base such decisions. We are surely better off with the fullest and most accurate understanding of social causal connections that we can obtain using sound methods, whatever the uncertainties besetting those results. But the soundness of our methods is important, for we may be worse off if we give our trust to falsehoods because they are produced by "scientific" methods that are in fact unsound.

The evidence that exposure to media violence causes later aggression rests largely on observational (nonexperimental) studies demonstrating a statistical association between media exposure and aggression outcomes (e.g., Huesmann, Moise-Titus, Podolski, & Eron, 2003; Johnson, Cohen, Smailes, Kasen, & Brook, 2002; Stacy, Smith, & Donnerstein, 1998), augmented by some short-term experimental studies. Inferring causal information from observational designs generally requires strong assumptions, and such inferences are especially vulnerable to flaws in data collection or analysis. Even in the face of these vulnerabilities, however, we can often learn information of substantial value to individuals and policy makers from a body of well-conducted observational studies. In this article, we review some of the difficulties in causal inference from observational, longitudinal data; examine the extent to which these seem relevant to the empirical work on exposure to televised violence published to date; and present a reanalysis of data from one especially influential study to address one of the more serious limitations of existing analyses. We conclude that the data we consider gives evidence that there is likely, although not certainly, a causal connection between exposure to televised violence and adult aggression.

Inferences to a causal relation between exposure to TV violence and adult aggression from observational data are frequently criticized on the grounds that correlation does not imply causation. This claim is certainly true, but nonetheless, the pattern of statistical associations observed among sets of variables in a sample can sometimes be used to estimate causal relations, and techniques to identify causal effects from statistical associations are the bread and butter of quantitative social science. Because the

task of estimating the effect of exposure to TV violence on aggression is beset by difficulties similar to those in much observational research, we begin by itemizing some of the factors that can, and sometimes do, produce incorrect estimates of a causal relation between an exposure and an outcome (some of these are discussed in more detail in later sections of the paper):

1. There may exist a plethora of alternative causal models that *might* explain the data consistently with prior knowledge of domain experts, often far too many models to survey or test individually. Standard hypothesis testing compares only two models (a null and one alternative). Hence, the conclusion that a model passes a significance test should provide little reason to prefer this model over any number of other, unconsidered alternatives.
2. Sampling that is not independent and identically distributed (i.i.d.), either because of temporal relations or individual-to-individual interactions or because of mixed samples from populations with distinct causal relations; nonrepresentative samples can lead to incorrect estimation of the underlying causal relations.
3. Sample bias affecting both nonexperimental and experimental studies, in which values of variables for an individual influence whether that person appears in the sample. Sampling that is dependent on the effect will bias the estimate of linear dependence on the cause, and sampling that is influenced by dependency on any two variables will generally make these variables dependent in the sample even if they have no causal connection.
4. Sample data that are collected in multiple disparate studies with distinct but overlapping sets of variables. Related studies with slightly different variables are common enough in social science and make meta-analysis difficult.
5. Nonlinear dependencies among continuous variables. Methods of estimating effects appropriate for linear systems and Gaussian, or non-Gaussian distributions may be badly biased when dependencies are nonlinear, and the bias may affect many aspects of model specification.
6. Confounding of associations among recorded variables by variation of unrecorded, unknown common causes. A notorious source of erroneous causal conclusions from observational data can arise whenever the variables statistically controlled for are imprecisely measured or otherwise inadequate; and, if the variables controlled for are actually effects of the exposure or outcome, their statistical control will further bias causal estimates.[1]
7. Misleading measures of uncertainty for estimates of causal strength. In a linear model, one measure of uncertainty is the confidence interval, but confidence intervals are not available for causal strength when latent variable alternative models are allowed. Non-i.i.d. sampling, if unaccounted for, usually leads to overly narrow confidence intervals.

These difficulties result from the mathematical structure of the inference problem and do not include the difficulties specific to each study, such as obtaining valid measures of the outcome (effect) under study. When the effect is described in colloquial terms, for example, “aggression,” it may include different features in

different studies and may not be a feature that can be reliably projected to a more general population.

Most, but not all, of the above issues arise for studies of the connection between child and adult television watching and aggression in the same adult. Some of the difficulties mentioned above can be met by resort to Bayesian statistics, where model uncertainties are explicitly part of the mathematical formalism. Advocates of consilience or "convergence" arguments infer causal connections from significant associations found in many different studies. These arguments are perhaps best viewed as informal Bayesian arguments for a qualitative conclusion: The consistent results in diverse studies are unlikely if no causal connection exists, but probable if such connection does exist. But formalizing Bayesian arguments and statistics so that they are a reasonable idealization of anyone's (let alone almost everyone's) degrees of belief is not easy, and the literature tends to veer between idealizations that are unlikely to approximate anyone's actual betting odds, and completely informal consilience arguments. A correct assessment of the arguments for a causal connection between exposure to televised violence and aggression requires more than nonclassical statistics.

Minimally, causal inferences from observational data must proceed from some understanding of the extent to which the data underdetermine the true causal structure relating measured variables. That is partly a matter of the range of causal structures with which the data are consistent. Specifying that range requires effective search over as many potential causal structures as possible to identify those potential explanations that predict the data with some sufficiently high probability.

This article is organized as follows: The next section introduces some semiformal background and develops some of the assumptions needed for causal inference from observational data, summarizing two decades of work in philosophy, statistics, and computer science. We then note the extent to which we may—and may not—reasonably regard the assumptions required for causal inference as being satisfied in the case of media exposure to violence and subsequent aggression. The next few sections discuss some of the flaws in various recent efforts to establish that exposure to television violence increases aggression, but note that for the most part, these flaws are not fatal. Finally, we apply search procedures resulting from the above-mentioned progress to data from one of these studies.

Background

Current discussions of causal inference have two paradigms, the "counterfactual" framework predominant in statistics and the graphical causal modeling framework predominant in computer science. The formalisms employed by the two frameworks are closely related, and there have been various arguments that they are equivalent. Whether equivalent or not, we will use the graphical model formalism, in part because

it lends itself to search methods and because graphical rather than algebraic representations tend to be more accessible. Readers are referred to Spirtes, Glymour, and Scheines (1993, 2000) and Pearl (2000) for more formal treatments of the general issues.

Among the several empirical techniques used to draw causal inferences in the graphical causal model framework, one important technique is to estimate a family of conditional probability distributions from an observed sample data set. Given a set of measured variables, say Drinking, Studying, and Passing Exam, the job is to characterize the joint probability distribution or density over these variables given sample frequencies as data. Given a hypothetical intervention that randomizes a variable X, we suppose there is a conditional distribution for all other variables in the system. Relations among such distributions are represented in intuitive fashion by directed acyclic graphs (DAGs). In a DAG, directed edges (arrows) represent direct causal effects of variables—which are represented as vertices—upon one another (as in Figures 1 through 3). Each DAG entails a set of conditional independence relations, and the search procedures test those relations in the sample data, eliminating possible causal connections if a hypothetical independence relation is rejected by a test. Algorithmic details and proofs of correctness are given in Spirtes et al. (1993, 2000). Alternative search methods may be Bayesian, or may use scores for causal models other than their posterior probability distribution, or may use the sample to test for constraints that may be implied by various models. Prior causal information, as from temporal relations among measurements, may be used to constrain permissible models, although it is not necessary.

Once a causal DAG or DAGs are identified that best explain the conditional independence relations among the variables, each DAG may be turned into a specific linear or nonlinear model, according to the distribution family (e.g., Normal or Multinomial) used in the search. The parameters of such models (e.g., linear coefficients) can then be estimated from the sample data, and the models can be subject to further testing and use in forecasting or predicting effects of interventions.

The graphical model framework uses two primitive notions: (a) variable X is a direct cause of variable Y relative to set $\boldsymbol{V}$ of variables for unit u; and (b) f is a probability distribution over joint assignments of values to variables in $\boldsymbol{V}$.[2] Effective search over possible models for those that explain the data depends on the relationship between these fundamental concepts. Given a set of variables $\boldsymbol{V}$, the notion of direct cause posits a binary relation. We will assume it is asymmetric: for all X, Y, at most one of X, Y is a direct cause of the other (i.e., if X causes Y, Y cannot simultaneously cause X). That is not a plausible assumption in all cases (e.g., not for motion of a bicycle pedal and the motion of the rear wheel of the bicycle), but it is appropriate for the cases under discussion. Binary asymmetric relations on a collection of objects determine a DAG; in this case, the objects are variables for a unit. So we assume that with each unit u to be considered there is a DAG, $G(u)$,whose vertices are variables and whose directed edges represent the proposition that the tail vertex is a direct cause of the head vertex.

Figure 1
Directed Acyclic Graph (DAG) Depicting Hypothetical Causal Structure Relating Variables

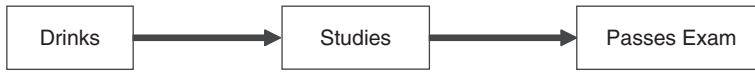


Figure 2
Directed Acyclic Graph (DAG) Depicting an Alternative Hypothetical Causal Structure Relating Variables

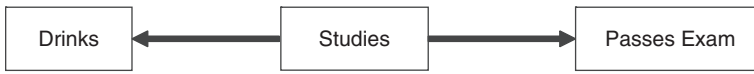


Figure 3
Directed Acyclic Graph (DAG) Depicting a Third Possible Hypothetical Causal Structure Relating Variables

Three illustrative DAGs, or causal graphs, are given above. In Figure 1, Drinks is a direct cause of Studies, which is a direct cause of Passes Exam; Drinks is *not* a direct cause of Passes Exam, but rather an indirect cause. In Figure 2, Studies is a direct cause of Drinks and also a direct cause of Passes Exam, but Drinks is not a cause of Passes Exam at all. And in Figure 3, Drinks and Studies are both direct causes of Passes Exam, but Drinks is not a cause, direct or indirect, of Studies, and conversely Studies is not a cause of Drinks. Each such graph represents possible causal structures (see Figures 1 through 3).

The set of all possible causal structures relating a given collection of variables is easily specified using the graphical framework. For example, we have here three variables giving us six possible directed edges yielding 2^6 or 64 possible graphs relating Drinks, Studies, and Passes Exam. If, in keeping with the assumption of asymmetry, we forbid closed causal pathways from a variable back to inself, this number is reduced: We have 3^{3-2} or 5 possible graphs. If unmeasured common causes are permitted, the number of possibilities expands by a factor of 8, because there are 8 possible ways to introduce confounding variables: one between each pair of measured variables, one between each of any two pairs (3 ways), one between any one pair (3 ways) and none.

The causal assumptions represented in the DAGs mathematically imply specific sets of statistical associations among the variables. This fact—that causal relations induce statistical associations—is the key to drawing causal inferences from patterns of statistical relations. The rules for finding the statistical associations implied by any set of causal relations are simple; we introduce some terminology before giving the rules.

A directed path in G from X to Y is a sequence of variables, the first of which is X and the last of which is Y, such that every variable in the sequence is a direct cause of the next variable in the sequence. So in Figure 1, the sequence Drinks, Studies, Passes Exam is a directed path, but the same sequence is not a directed path in Figure 2. We say that a subset W of V is causally sufficient provided that for every pair of variables X, Y in W all of their confounding common causes are also in W. For example, if the DAG in Figure 2 is the correct causal structure over Drinks, Studies, and Passes Exam, then the set {Drinks, Passes Exam} is not causally sufficient, because a common cause, Studies, is omitted from the set. However, the set {Drinks, Studies} is causally sufficient for the causal structure represented by the DAG in Figure 2.

Testing for the consistency of a causal model whose structure is represented by a directed graph requires some assumptions. The first is the standard assumption that one's data set is representative of the true probability distribution over measured variables. The assumption may be in error, and it is in various ways subject to test, but it is also essential to any causal inference, whether from experimental or from observational data. The second assumption is the Causal Markov Condition (CMC). The CMC constrains the probability distributions that are imposed on variables in a causal hypothesis. The intuition for CMC is simple: If the causal structure is $Z \rightarrow X \rightarrow Y$, then changing the value of Z will not affect Y if X is forced to stay constant, because Z only influences Y through X. More formally, CMC says that in a causally sufficient system V of variables (i.e., a set of variables that includes all common causes on its members), conditional on assignments of values to its direct causes, each variable X is independent of any other subset of V not containing effects of X. Assuming sample data are representative of the true probability distribution over measured variables, this constraint will hold in the sample distribution no less than in the true probability distribution.

To see how the CMC permits tests for the consistency of model with sample data, consider again our three illustrative graphs. Given the CMC, the DAG in Figure 1 implies that Drinks is associated with Studies, Drinks is associated with Passes Exam, and Studies is associated with Passes Exam. It also implies that if one conditions on, that is, statistically controls for, Studies, then Drinks and Passes Exam will be independent, but that if one conditions on Passes Exam, Drinks and Studies will remain associated. If, in the sample data, these associations and independencies do not hold, then Figure 1 cannot be a correct model of the causal structure relating Drinks, Studies, and Passes Exam (unless, of course, one's sample data do not represent the true distribution over the three variables).

Effective searches are limited to models that satisfy a further constraint, called the Faithfulness Condition, and it is essentially the converse of the CMC. It says that probabilistic independencies and conditional independencies (again: nonassociations between variables when the data are stratified by potential confounders) arise only from the CMC applied to the graph of causal relations. In linear systems, this amounts to assuming that measured variables are not deterministically related and the effects of multiple causal pathways connecting two variables do not perfectly cancel one another.

To see which causal models are omitted by the Faithfulness Condition, consider a fourth possible causal structure relating Drinks, Studies, and Passes Exam, this time including edge loadings (e.g., standardized regression coefficients). If the causal structure over Drinks, Studies, and Passes Exam is as in Figure 4, and $\alpha = -1(\beta\gamma)$, then Drinks will be statistically independent of Passes Exam, even though Drinks is a cause of Passes Exam. The Faithfulness Condition rules such situations out: To assume faithfulness is to assume that when such structures arise, $\alpha \neq -1(\beta\gamma)$. Given the Faithfulness Condition, a graphical structure implies quite a number of useful facts about the (in)dependencies and conditional (in)dependencies that must hold between measured variables if the graph represents the true causal structure over those variables. Suppose, for example, we observe the following in our sample data:

1. Passes Exam is probabilistically dependent on both Drinks and Studies,
2. Drinks is probabilistically independent of Studies, and
3. Drinks and Studies are probabilistically dependent when controlling for Passes Exam.

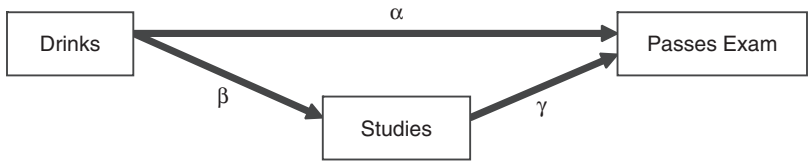
The causal structures represented by the DAGs in Figures 1 and 2 are inconsistent with these findings; under the assumptions expressed in both Figure 1 and Figure 2, we would expect Drinks and Studies to be statistically associated, contrary to finding 2) above. Among the three DAGs in Figures 1 through 3, the DAG in Figure 3 uniquely implies the qualitative statistical relations between measured variables. Conversely, suppose we examine the sample data and find, that

1. Drinks, Studies, and Passes Exam are all associated,
2. Drinks remains associated with Studies conditional on Passes exam, and
3. Drinks and Passes Exam are independent conditional on Studies.

Then Figure 3 cannot be correct because it is inconsistent with findings 1) and 3) above; but both Figure 1 and Figure 2 imply all of these dependence and conditional independence relations, and so either might correctly represent the true causal structure.

The rules for reading off the dependence and conditional dependence relations implied by a causal structure are as follows; they come from the d-separation theorem (Pearl, 1988). Any two variables *X* and *Y* will be associated conditional on a set *V* of variables, with *X* and *Y* not in *V*, if and only if there is an active path between *X* and *Y*. A path is active if every node in the path (other than *X* and *Y*) is active.

Figure 4
Hypothetical Directed Acyclic Graph (DAG), With Loadings



A node (variable) *Z* is active on a path if (a) it is a "mediator," that is, on the path *Z* has one edge directed in and one edge directed out (← *Z* ← or → *Z* →), and *Z* is not in *V*; (b) it is a "common cause," that is, has in the path two edges directed out (← *Z* →) and is not in *V*; or (c) *Z* is a "collider," that is, has in the path two edges directed in (→ *Z* ←) and either is in *V* or has some effect, direct or indirect, which is in *V*. For small graphs that are not dense (few edges), one can determine the (in)dependencies and conditional (in)dependencies implied by a graph; even for large graphs with dense numbers of edges, it is possible to compute the predictions the graph implies algorithmically. Doing so, and checking predictions against measured (in)dependencies and conditional (in)dependencies, provides a way to search over a very large range of possible causal models for those that are consistent with the data.

The general idea, then, is this: Given assumptions about the relation between causal structure and probability distributions (the Faithfulness Condition and CMC) and between probability distributions and sample distributions (that sample error is negligible and that the sample size is high enough to permit detection of the relevant [in]dependence and conditional [in]dependence relations), possible causal structures can be tested against data in a purely qualitative manner. Each causal structure will imply a set of probabilistic (in)dependencies and conditional (in)dependencies. If these do not hold in the data, then either the sample data do not represent the true population distribution or the causal model is not correct. Because the graphical representation of the causal structure lends itself to computational methods, it is possible to search over possible causal structures relating a set of measured variables to identify those, and only those, that predict the measured dependencies and independencies.

Of course, quite apart from worries about sample error, the set of causal structures consistent with measured dependencies is likely to be large, although much smaller than the set of all possible structures. And many such structures, while qualitatively consistent with the data, will be quantitatively inconsistent, in that there are no fully specified models with that structure that pass, for example, a chi-square test on the data. Nonetheless, searching for causal structures qualitatively consistent with the data is useful for several reasons. Two are of particular importance here. First, it may be that all the causal structures consistent with the data share one or more particular causal relations, for example, in every such graph there is a direct path from Drinks

to Passes Exam. In this case, one has very good reason to say that Drinks really does cause Passes Exam. Second, even if there are no such common features, one nonetheless has, in the search results, both a characterization of the extent to which the data underdetermine the causal truth and also of the range of models that must be checked quantitatively against the data. The assumptions connecting structure to probability distributions are absolutely essential, and we discuss them further in the next section.

Automated, algorithm-based search procedures are entirely consistent with conventional hypothesis-testing approaches but employ two important innovations. The first innovation is the idea of searching over a large set of possible causal structures. Conventional hypothesis-testing approaches begin by specifying two possible causal models, one in which the exposure of interest does not affect the outcome (the null model) and another in which it does (the alternative model). The association between the exposure and the outcome is then examined to see whether it is inconsistent with the null model. Most research studies also examine results after adjusting for a set of covariates believed to potentially influence the exposure and the outcome. Even such adjusted models implicitly "test" only a very limited number of causal structures, typically structures in which the measured covariates are causally unrelated to one another and are common causes of the exposure and the outcome of interest. Under a tremendous range of alternative causal structures, including causal structures with no omitted common causes, the adjusted models would fail to identify the causal effect of the exposure because the implicitly assumed structural relations among the measured variables are incorrect.

The second, related, innovation arises from integrating the d-separation rules with the Faithfulness Condition. This renders the statistical dependencies and independencies observed in sample data much more informative about the causal structures. For example, under the d-separation rules, conditioning on a common effect of two independent variables induces a statistical association between them. This fact is exploited in the automated discovery algorithms but is almost never used to select causal models in conventional hypothesis testing. Thus, if variables X and Y are statistically independent before conditioning on any other covariates but become statistically associated when conditioned on a third variable Z, under faithfulness this implies that X and Y both affect Z (or that causes of X and Y affect Z). This new sort of information is rarely used in conventional approaches and indeed its use in automated algorithms is controversial, in part due to its reliance on faithfulness. Without the faithfulness assumption, this combination of observations (X and Y marginally independent but statistically associated conditional on Z) is consistent with a number of other causal structures that include perfectly counterbalanced paths. We discuss this controversy in more detail in a later section. These concerns are not trivial, but nor should they preclude us from using the data to gather as much information as possible about the underlying causal structure. These computerized methods do not provide a royal road to the true theory, but, by

searching automatically over an astronomical space of alternative structures and finding alternatives that explain the data, they certainly do provide a royal road to caution, and to localization of uncertainties, in causal inference.[3]

Assumptions and Constraints

From the graphical causal modeling perspective, there are three crucial conditions that must be met if causal discovery is to be possible. First, measured associations must not arise simply as a result of sampling error, and second and third the CMC and Faithfulness Condition must be satisfied. Several further assumptions and sources of further information can aid causal discovery, among them information about the form of the probability distribution over measured variables and a priori constraints on causal relations. In this section, we review the warrant for these conditions with respect to data on exposure to televised violence and aggression.

Suppose we observe statistical associations between measured variables in a sample drawn from a population using a partially specified selection rule (we know some, but not all, the determinants of sample membership). What could account for these observed associations? Associations between two measured variables in a sample can be generated in a variety of ways:

1. one variable causing another,
2. common (confounding) causes,
3. sampling error, and
4. sampling bias.[4]

In all but the case of sampling error, and even in some special instances of this, the associations are induced by the causal structure relating variables in the population and sample membership. If associations found in one or several studies result from sampling error, then searches over alternative models for those that causally explain the measured associations will be fruitless or misleading: Automated search methods begin from the assumption that associations are generated by the causal structure among measured variables.

Consilience or "convergence" arguments point to agreements in sign and direction of effects in multiple observational studies and infer from this agreement that associated variables must be causally connected by some directed path (e.g., Geen & Donnerstein, 1998). Here is a formulation from Rubenstein (1982):

> Because many correlational studies consistently show associations between heavy TV consumption and sociopathic attitudes and behaviors, there is probably a cause-and-effect relation between viewing and attitude formation and behavior. . . . Most television researchers look at the totality of the evidence and conclude . . . that the convergence of most of these findings about televised violence and later aggressive behavior by the viewer supports the positive conclusion of a causal relationship.

As arguments for the claim that exposure causes aggression, these arguments are unsound: Potentially many possible models will explain the associations, and in only some of them will exposure be a cause, direct or indirect, of aggression. But as arguments against the explanation of measured associations by appeal to sampling error, they are reasonably good informal arguments. Absent evidence of publication bias, explanations of associations that appeal to sampling error become increasingly implausible as sample sizes, or the number of "convergent" studies, increase. Sample sizes in studies on the relation between exposure to television and aggression are not large; neither are they trivial, and there are lots of such studies (see Pecora, Murray, & Wartella, in press).

If associations are not produced by sampling error, the statistical associations (those measured in the sample) are generated by causal relations among the variables in the population. There are various ways one can attempt to discover these causal relations from information about associations in the sample. Typical practice is to develop a model "by hand," working from expert judgment and piecewise correlations and partial correlations between outcome and variables already included in the model. One then tests the resulting model or models as to their fit with the quantitative data. As discussed in the previous section, this provides no assurance that the model is correct, rather than some unexamined alternative also fitting the data. It is typically not obvious what conditional dependencies and independencies are implied by a given model, and any set of measured dependencies and independencies is often consistent with a large number of models. If complaints against "convergence" are more than an appeal to the possibility of sampling error, this is just the worry that drives them.

Herein lies one significant advantage of the graphical models framework. Using that framework one can search systematically for models that explain the observed dependencies and conditional independencies, and one can rule out alternative causal structures that do not explain the observed associations. The CMC and Faithfulness Condition warrant some discussion. The CMC, or one or another informal equivalent, is routinely assumed, either explicitly or implicitly. For example, injunctions to statistically control for confounding variables and to avoid controlling for mediating variables when estimating the total effect of one variable on another implicitly appeal to the CMC. Assumptions of faithfulness are typically implicit. For example, inferences from the statistical independence of measured variables to the claim that these variables do not cause one another rely on faithfulness. That, of course, does not mean the assumption is true in every case, and in practice, given finite sample sizes, near violations of faithfulness can make causal inference difficult.

Assuming only the CMC and Faithfulness Condition and i.i.d. sampling, information about the causal structure can be consistently estimated from the conditional independence relations among the measured variables. In practice, one has only a sample, not a probability distribution over the recorded variables, and therefore extra

information or assumptions about the family of probability distributions are necessary for inference; and extra information, as about the time order of variables, is useful in eliminating models that would otherwise need to be considered. But principled search methods—those that provably give correct information given the probability distribution on the measured variables and the assumption that the true explanation is somewhere in the space of alternatives that can be effectively searched—are available under a range of assumptions, including those most common in social statistics: for Gaussian and for multinomial distributions, for assumptions of no feedback and no latent common causes, for cases where feedback (i.e., nonrecursive) structure is to be estimated from equilibrium data, for cases in which it is assumed there is no feedback but there may be latent variables, and for multiple indicator models whose latent structure may be recursive or nonrecursive.

If either the CMC or the Faithfulness Condition fail, reliable causal inference from the observational data is not possible among Gaussian variables. There are possible choices of variables and populations in which either or both do fail. Failures of either, however, require very special arrangements. Assuming only the CMC and Faithfulness Condition, some, information about the causal structure can often be learned from the observed associations. Extra information of various kinds increases what can be learned. Knowledge about the form of the probability distribution over variables, for example, whether the relations are linear, is especially useful, for determining whether causal connections exist and estimating path coefficients. Linearity is often assumed and sometimes explicitly tested (e.g., Johnson et al., 2002). Prior causal information is perhaps even more useful for discovering causal connections and orienting the direction of identified causal relations. For example, the temporal order of the variables implies restrictions on what can cause what. Two recent studies (Huesmann et al., 2003; Johnson et al., 2002) employ longitudinal designs that provide such prior causal information.

Liabilities to Causal Inference: Misplaced Conditioning

As is generally recognized, while statistically controlling for potential confounders, one must not condition on mediating variables, for doing so will remove nonspurious associations. What is not generally recognized, or recognized only in a special case, is that conditioning on the common effect of an exposure and an outcome variable will induce an association between exposure and outcome. Sampling bias is a special case of the general phenomenon. If, for example, exposure to television and aggression are causes of membership in a measured sample, associations in the sample are conditional associations—associations conditional on an effect of both exposure and outcome, namely, sample membership.

The failure to recognize the generality of this constraint sometimes leads to conceptual, though not necessarily substantive, errors. It is well known that conditioning on an effect of an outcome variable will lead to a biased estimate of the regression coefficient for the outcome on a putative cause, and this is duly noted in typical discussions. It is commonly not noted that conditioning on an effect, such as sample membership, of both the putative cause and the target effect will produce an association between variables that have no causal connection whatsoever. So, for example, in their study, Huesmann et al. (2003) note that more aggressive children are less likely to be represented in their fourth wave follow-up than less aggressive children. They claim, as is perfectly standard, that this differential representation in the follow-up sample may bias the estimate of the influence of exposure to television on aggression downward but is unlikely to lead to an incorrect inference that exposure to television causes aggression. But if adult aggression and childhood exposure to television are both causes of sample membership, exposure and aggression will be associated in the data, whether or not exposure causes aggression. We have no very good reason to think that this sort of sampling bias is in fact a problem here or in other studies of television exposure and aggression. Neither are we sure it is not, and we note that it is crucial to avoid conditioning on any common effects of exposure and outcome, whether explicitly in one's analysis or implicitly in constructing a sample.

Longitudinal Studies

Longitudinal studies have two important advantages over cross-sectional studies. First, longitudinal designs allow us to clearly establish that the exposure in question temporally precedes the outcome and thereby afford prior causal knowledge. Second, longitudinal designs allow us to examine whether the exposure predicts changes in the outcome. In a cross-sectional study, we may be concerned that both the exposure and the outcome are determined by an earlier (unmeasured) value of the outcome. We can rule out such explanations if we observe that the baseline value of the exposure predicts changes in the outcome variable over follow-up, unless we also believe that earlier values of the outcome influence the evolution of the outcome over time.[5] Similarly, it may be useful to test whether changes in the exposure predict changes in the outcome. Such a model can potentially rule out confounding by unobserved common causes of the exposure and the outcome, provided those unobserved confounders have time-constant effects (for a discussion of using panel data to overcome omitted variable bias, see chapter 10 of Wooldridge, 2002). The idea is that measured associations between exposure and outcome cannot be explained by a confounding time-invariant characteristic if changes in the exposure of interest predict changes in the outcome. For example, although a cross-sectional correlation between TV watching and aggressiveness might be explained by parent's education influencing both factors, this explanation is suspect—though still possible[6]—if

changes in TV watching predict changes in aggressiveness. To rule out unobserved confounders in this way, we must assume that the confounders have time-invariant effects, which may not appear to be a credible assumption in many situations, for example, when there is a natural developmental process.

Both the Johnson et al. (2002) and the Huesmann et al. (2003) studies can claim the first advantage: Their measures of exposure to television are temporally prior to their measures of aggression. However, neither study can make a compelling claim to the second sort of advantage, because neither study is based on a credible measure of change in the outcome of interest. Johnson et al. present a figure showing television viewing at age 14 and aggressive acts at ages 16 or 22, stratified by any prior violence and sex. But prior aggression is a continuous quantitative phenomenon, and dichotomization in this way leaves open the possibility of residual confounding in the group with some prior aggression. The group with a history of aggression probably includes a range of children with very mild aggressive acts to children with a history of frequent or severe aggression. In the subgroup with no prior aggression, the relation between TV viewing and aggression is weaker. A similar problem will influence the covariate adjusted logistic regression models, if the measure of prior aggression is similarly imprecise.

Huesmann et al. (2003) seem to suggest that they have partialed out baseline aggression, but this is almost certainly not the case. Huesmann et al. use the follow-up (Time 2) measure of the outcome variable as the dependent variable and adjust for the baseline measure. In general, such baseline adjusted models are not a reliable method for demonstrating the effect of an exposure on change. If the true effect of exposure on change is zero but there is a nonzero statistical association between exposure and the baseline value of the outcome, the effect estimated in a baseline adjusted model will generally be biased toward the nonzero baseline association when there is unreliability or instability in the baseline outcome measure (Yanez, Kronmal, & Shemanski, 1998; Yanez, Kronmal, Shemanski, & Psaty, 2002). If the baseline measure is unstable, an unbiased estimate of the effect of exposure on change can often be obtained by using the difference score (value of the outcome at follow-up minus value of the outcome at baseline) as the dependent variable (provided the regression model is not adjusted for the baseline measure of the outcome). However, Huesmann et al. did not adopt this approach, possibly because their two measures were on very different scales. To the extent that their baseline aggression measure is unstable or unreliable, their analysis is subject to the same biases as a cross-sectional study.

What Is Being Measured?

Huesmann et al. (2003) provide a very high estimate of the reliability of their aggression measure: 0.91 for 1 month test–retest reliability. Interestingly, the aggression

measures at baseline and follow-up are very weakly correlated: 0.17 for females and 0.20 for males (Huesmann et al., 2003, Tables 6 and 7). In fact, for females, childhood TV viewing is a more powerful predictor of adult aggressiveness than is childhood aggression. The low correlation between the Huesmann et al. aggression measures seems to be in contrast to published estimates of stability of measures of aggressiveness from childhood to early adulthood, which are typically in the range of 0.6 to 0.8 (Loeber & Hay, 1997). This suggests that either the Huesmann et al. variables are not measuring the same construct at the two time periods or that they are not measuring aggression as this is generally understood in the literature. One might argue that if the former, a major advantage of the longitudinal design is lost. If the latter, the resulting knowledge of causal structure will not in fact be useful for preventing aggression. But neither horn of the dilemma is disastrous. First, if the Huesmann et al. study does not have all the advantages of a longitudinal study, it nonetheless does allow inference to some features of the causal structure within which adult aggression is embedded. Second, despite the surprisingly low stability of the Huesmann et al. aggression measures, we note that there is substantial value in Huesmann's strategy of augmenting objective measures of violent behavior with subjective measures such as the conflict tactics scale (CTS), as H the CTS provides a measure of violent behaviors of substantial importance that are unlikely to be recorded as crimes and therefore unlikely to be represented in objective measures of aggression such as criminal records.

Instrumental Variables

Instrumental variables (IV) analysis provides a second procedure for identifying causal effects from observational data (Angrist & Krueger, 2001; Greenland, 2000). The central idea of an IV analysis is to find a variable that is associated with the exposure of interest (TV viewing) but would have no other causal link to the outcome (i.e., it does not itself directly affect the outcome or share an unmeasured common cause with the outcome and is not itself an effect of the outcome). The most promising instruments arise from natural experiments, for example, the Williams (1986) study of Canadian towns before and after the availability of TV signals. The challenge of implementing IV approaches is to find good instruments: natural experiments for TV viewing are not common. Luketsch (1989) adopted the questionable approach of instrumenting viewing violence in one medium (TV) with measures of violence exposure from another medium (videos). This violates the assumptions for a valid instrument, because exposure to violence through either medium could be either a direct cause or effect of aggressive behavior. IV analyses do little to strengthen causal inference if the instruments are of questionable validity. Regrettably, social survey data are rarely chosen with a plan for measuring IVs for causal relations that are of interest but that may be confounded. This is one place where "substantive knowledge," which social scientists often claim is essential to model

specification, would be of real value. Plausible statistical criteria for instruments are available, but rarely applicable. For example, if the estimated effect of X on Y is found to be the same with several, diverse possible IVs that would argue that each such variable is actually an instrument, that is, satisfies all of the necessary conditions given parenthetically above.

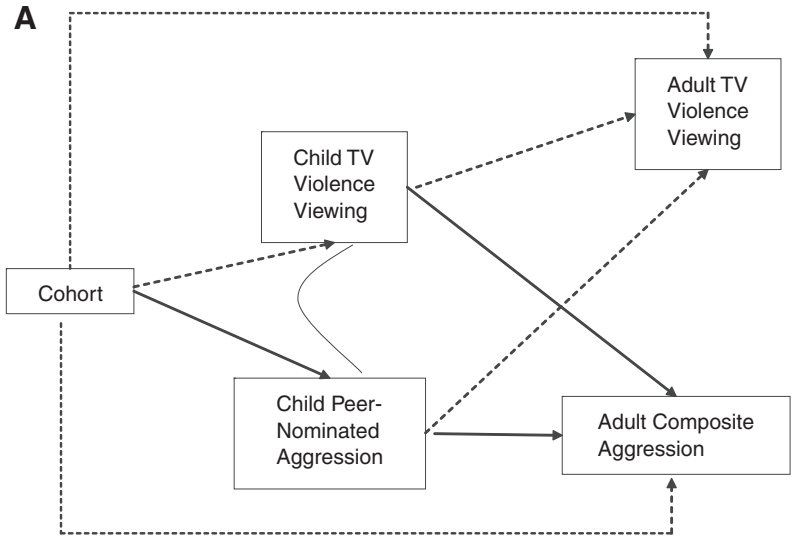
Search

The Huesmann et al. (2003) study, but not the Johnson et al. (2002) study, make explicit causal claims. That is perhaps to the credit of the former rather than the latter, because the interest in both studies is of course in causation. Huesmann et al. provide a pair of causal models for each gender and estimates of the path coefficients for those models. They do not, however, report the results of any search over alternative models, and any conclusions about the effects of childhood television exposure on adult aggression should wait at least upon the results of such a search. Huesmann and colleagues kindly granted us access to the correlations among variables in their models, allowing us to report the results of such a search.[7] The searches readily find models that fit the Huesmann et al. data well and that share common features with their models, provided it is assumed there are no unmeasured common causes at work. When that assumption is given up, however, the picture changes.

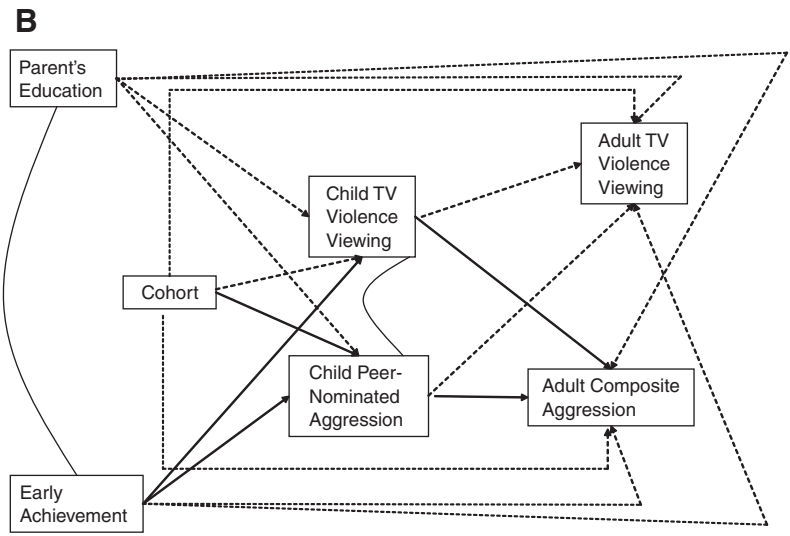
Huesmann et al. (2003) considered the correlations among 16 variables, 11 of which measured features of participants' parents (parent variables), 3 of which measured features of participants as children (child variables), and 2 variables that measured features of participants as adults (adult variables). They construct two structural models for each subpopulation; in every model, a measure of childhood exposure to violent television is a cause of a measure of adult aggression (Child TV Violence Viewing and Adult Composite Aggression, respectively). These are given without the edge loadings in the originals; coefficients for dashed edges are nonsignificant (see Figures 5 and 6). The first of each pair of models explicitly introduces a single confounder (cohort) and also includes an unexplained association (a correlated error) between Child TV Violence Viewing and a measure of childhood aggression (Child Peer-Nominated Aggression). In the second model for each subpopulation (males/females), two further confounders, Parental Education and Childhood Achievement Score, are introduced, again with an unexplained association between them. Most models fit the data well or reasonably well, with p values ranging from .99 to .43. Only one model, for females, does not fit particularly well, with $p = .04$.

Using automated procedures (the FCI and GES algorithms available at http://www.phil.cmu.edu/projects/tetrad_download/), we searched for classes of models that qualitatively explained the measured associations and conditional independencies between variables and fit the data quantitatively as assessed by chi-square tests. The search procedures in question allow the investigator to place prior constraints

Figure 5

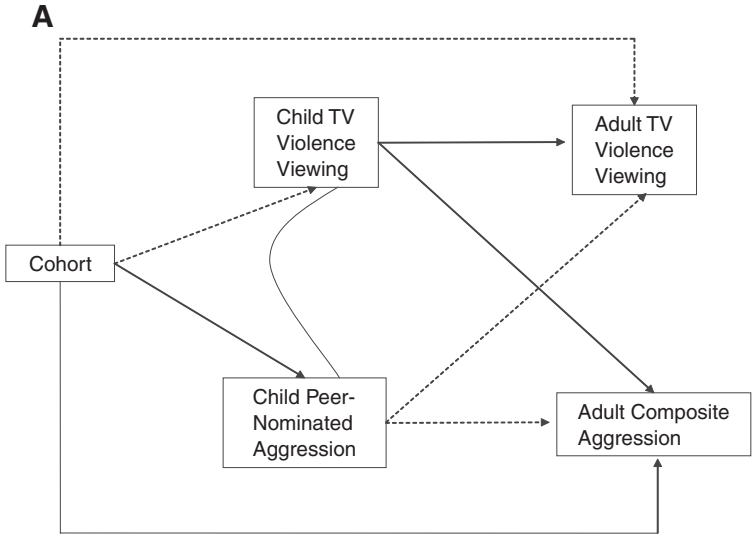


Note: Regression Model 1 for Males, from Huesmann, Moise-Titus, Podolski, and Eron (2003). Edge loadings omitted; solid edges represent significant associations, dotted edges nonsignificant associations; $p = .99$.

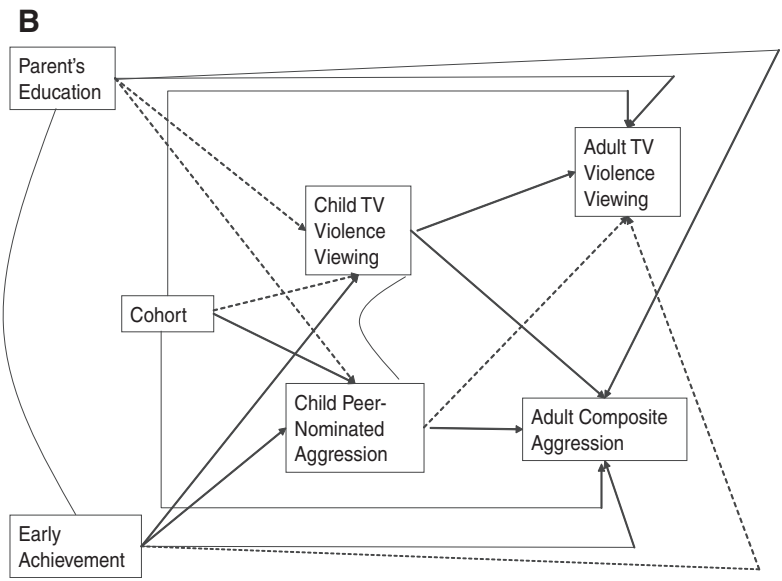


Note: Regression model 2 for Males, from Huesmann, Moise-Titus, Podolski, and Eron (2003). Edge loadings omitted; solid edges represent significant associations, dotted edges nonsignificant associations; $p = .94$.

Figure 6



Note: Regression Model 1 for Females, from Huesmann, Moise-Titus, Podolski, and Eron (2003). Edge loadings omitted; solid edges represent significant associations, dotted edges nonsignificant associations; $p = .04$.



Note: Regression Model 2 for Females, from Huesmann, Moise-Titus, Podolski, and Eron (2003). Edge loadings omitted; solid edges represent significant associations, dotted edges nonsignificant associations; $p = .43$.

on the possible causal structure and then search over possible structures consistent with those constraints to find those that fit the data qualitatively, those that predict the measured dependencies and conditional independencies given the CMC and Faithfulness Condition.

The reliability of search procedures based on FCI and related algorithms has generated some controversy (Robins, Scheines, Spirtes, & Wasserman, 2003; Zhang & Spirtes, 2003). Some researchers argue that the Faithfulness assumption is too strong, because opposing causal effects may frequently perfectly, or nearly perfectly, counterbalance in observed data. That is a correct point about the possible causal explanations that may be missed by computerized searches. We argue that it is nonetheless useful to search over the vast set of alternative causal explanations in which opposing causal effects do not counterbalance. Such a search seems especially valuable when, as in the circumstance here, previously established statistical associations do not uniquely determine the causal relations among the variables. The time order of measurements or other known features of the relationships among the variables can be explicitly entered into the search procedure to restrict the set of possible causal structures. Under those conditions, any model produced by automated search is prima facie as viable as one produced by a person without automated search. The superiority of models produced without a comprehensive search across causal structures (e.g., conventional regression models) to models produced by search under the Faithfulness assumption must be justified by substantive knowledge or by better fit. Were the full data available, the Faithfulness assumption might well be dispensable.

We therefore report here what can be learned using the FCI and GES algorithms under the Faithfulness Condition. Our approach is consistent with using theoretical knowledge to rule out some causal structures, or to require specific causal relations, provided that such knowledge claims are well-founded. In this case, however, we believe theoretical understanding of the relevant social processes is uncertain, so we rely only on basic assumptions such as temporal order, as well as the assumptions of Faithfulness and the CMC, to generate the results that follow.

Appropriate search procedures depend on the assumptions one is willing to make about the causal completeness of the measured variables. That is, one should use different procedures depending on whether one is willing to assume there are no unmeasured confounders. In effect, Huesmann et al. (2003) assume that there are no unmeasured confounders of Child TV Violence Viewing and Adult Composite Aggression. Following their lead, we employed the GES algorithm, a Bayesian search procedure whose consistency assumes the observed variables are causally sufficient. We also employed the FCI algorithm, which makes no such assumption, to do the same thing. The output of both procedures is not a DAG but rather a graphical pattern—edges may be directed or undirected—representing a Markov equivalence class of DAGs (i.e., the class of graphs that explain the measured dependencies and conditional independencies as required by the CMC and Faithfulness Condition).

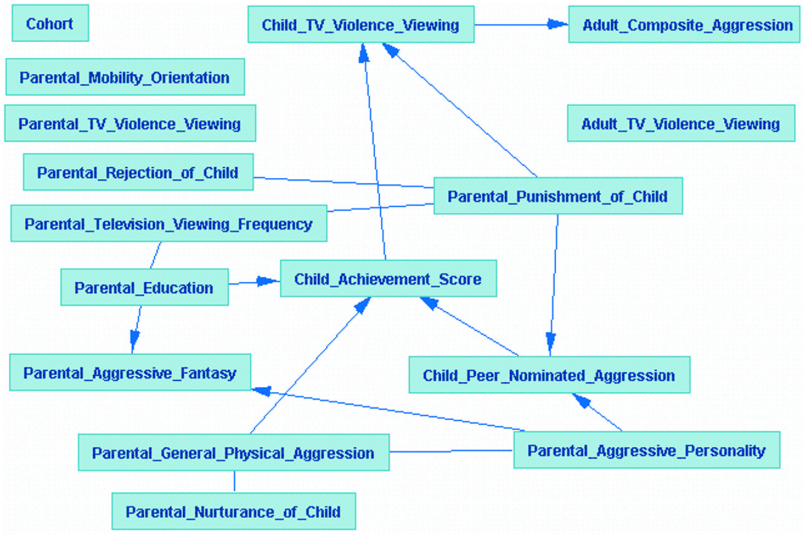
Directed edges (arrows) represent a causal relation between the connected variables, with the arrow pointing to the effect. In the GES output, undirected edges represent the presence either of an unmeasured common cause or a direct causal relation between connected edges. In the FCI output, bidirected edges represent the presence of an unmeasured common cause, partially oriented edges (with circle at one end and an arrowhead the other) imply that either the first causes the second or they share an unmeasured common cause, and unoriented (circle at both ends) edges imply that either the variables share an unmeasured common cause or one variable causes another and the algorithm cannot tell which.[8]

We then estimated and tested models derived by hand from the equivalence classes of models the search procedures returned. We first replaced undirected edges with directed edges or latent variables (when required by the pattern), choosing the direction of edges arbitrarily, as long as this was consistent with the returned pattern. The result is a DAG in the equivalence class identified by the returned pattern. Each DAG represents a distinct causal structure, which we then parameterized as a linear model and estimated by maximum likelihood. Measured associations among the 16 variables varied in sample size (from 89 to 176) as a result of missing data. We ran the algorithms using sample sizes corresponding to the arithmetic mean of sample sizes among the measured correlations (132 for males, 153 for females).

GES returns the following pattern for males. The pattern specifies the Markov equivalence class of models that explain the measured correlations and conditional independencies: In all such models, Childhood TV Violence Viewing is a direct cause of Adult Composite Aggression. Specific models in the pattern can be found that fit the data with a p score of .9470. Assuming there are no unmeasured common causes (as GES does), there are models that fit the Huesmann et al. (2003) data quantitatively and explain the measured dependencies and conditional independencies, and in all of them, Child TV Violence Viewing is a cause of Adult Composite Aggression (see Figure 7).

But of course, the assumption that there are no unmeasured common causes is not sacrosanct. FCI employs no such assumption, and so is more robust against unmeasured confounders. FCI returned the pattern in Figure 8 at $\alpha = .05$. The relevant feature is the unoriented edge between Child TV Violence Viewing and Adult Composite Aggression; this feature is robust for α between .1 and .05, but not for α between .05 and .01.[9] In fact, Adult Composite Aggression cannot be a cause of Child TV Violence Viewing; hence, if the FCI output is correct, either Child TV Violence Viewing is a cause of Adult Composite Aggression, they share a common cause, or both (see Figure 8). Models in the pattern representing the first possibility can be found that fit the data, with p scores of .4046. But similarly, models representing the second possibility can be found, fitting the data with p scores of .2349. If one is unwilling to assume that there are no unmeasured common causes of Child TV Violence Viewing and Adult Composite Aggression, one cannot eliminate this possibility using the Huesmann et al. (2003) data. On the other hand, the data do

Figure 7
GES Output for Males From Huesmann Data



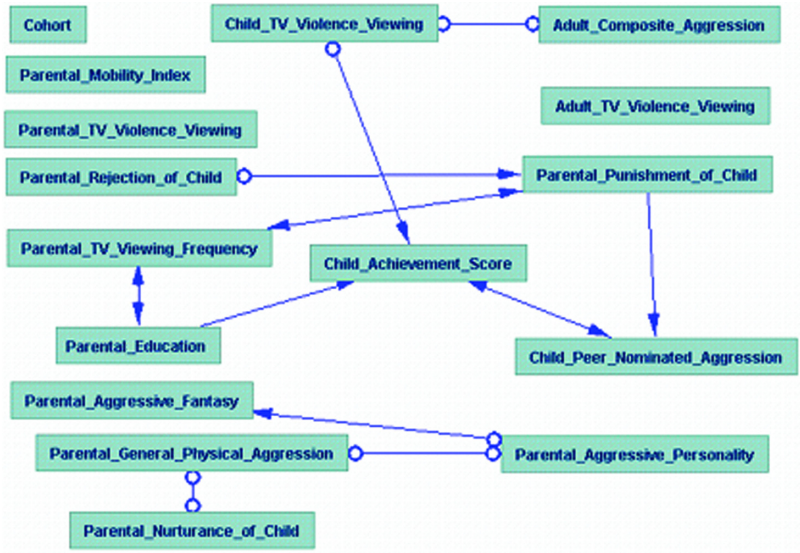
Note: Directed edges (arrows) represent a causal relation between the connected variables, with the arrow pointing to the effect; undirected edges represent the presence either of an unmeasured common cause or of a direct causal relation between connected edges.

support the claim that there really is some causal connection between childhood exposure and adult aggression: Assuming the measured correlations and conditional independencies really do hold in the population, either childhood exposure is a cause of adult aggression or they share some unmeasured common cause.

Matters are interestingly different with respect to females, for which GES returned the pattern (see Figure 9). Note that Child TV Violence Viewing is not a cause, direct or indirect, of Adult Composite Aggression. On the other hand, Adult TV Violence Viewing is. Moreover, models in the pattern can be found that fit the data quantitatively, with p scores of .4618. Assuming, again, that there are no unmeasured common causes, Adult TV Violence Viewing is a cause of Adult Composite Aggression in every model that explains the measured dependencies and conditional independencies and fits the Huesmann et al. (2003) data by a chi-square test.

For females, FCI returns the pattern in Figure 10 at $\alpha = .05$. This pattern requires that Adult Composite Aggression and Adult TV Violence Viewing share an unmeasured common cause, and includes models in which there is a further direct causal relation between Composite Aggression and Adult TV Violence Viewing (either variable

Figure 8
FCI Output for Males From Huesmann Data

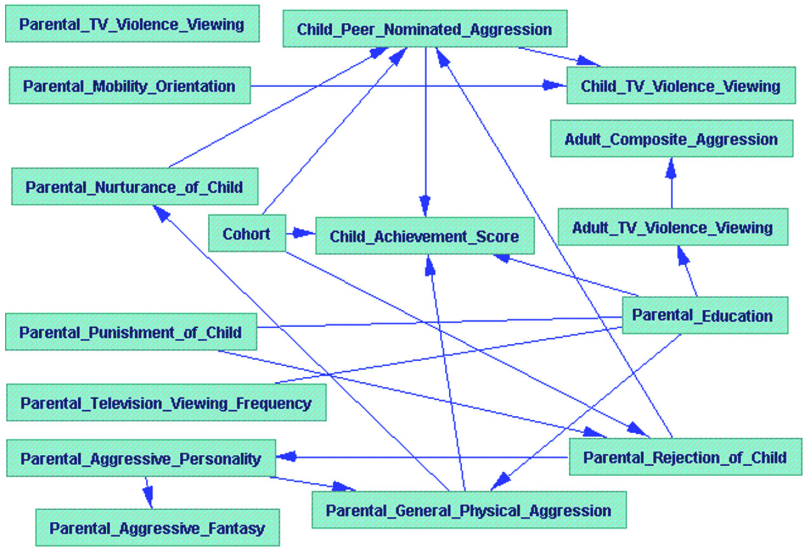


Note: Directed edges (arrows) represent a direct causal relation between the connected variables, with the arrow pointing to the effect; bidirected edges represent the presence of an unmeasured common cause; partially oriented edges (circle at one end and an arrowhead at the other) imply a direct causal relation or a shared unmeasured common cause; and unoriented (circle at both ends) edges imply that either the variables share an unmeasured common cause or one variable causes another but the algorithm cannot determine the direction of causation.

may consistently play the role of cause). The presence of an edge between these two variables, but not its orientation, is robust to changes in α between .1 and .015; the edge is lost at $\alpha = .01$.[10] Replacing the direct influence of Adult Television Watching on Adult Composite Aggression found by the GES search with a common cause of those two variables (in the form of a correlated error), as suggested by FCI, yields a model with $p = .268$. On the whole, then, the Huesmann et al. (2003) data show that among females, there is some causal connection (i.e., one causes the other or they share a common cause) between Adult TV Violence Viewing and Adult Composite Aggression. But unless one makes strong assumptions about unmeasured common causes, the data are inconclusive about the correct explanation of the association.

There is more to be learned from the female data. Neither GES nor FCI return patterns in which Child TV Violence Viewing is or could be a cause of adult aggression among females (this is especially interesting in light of the odds ratios Johnson et al., 2002, report for women). This feature is stable across all patterns

Figure 9
GES Output for Females From Huesmann Data

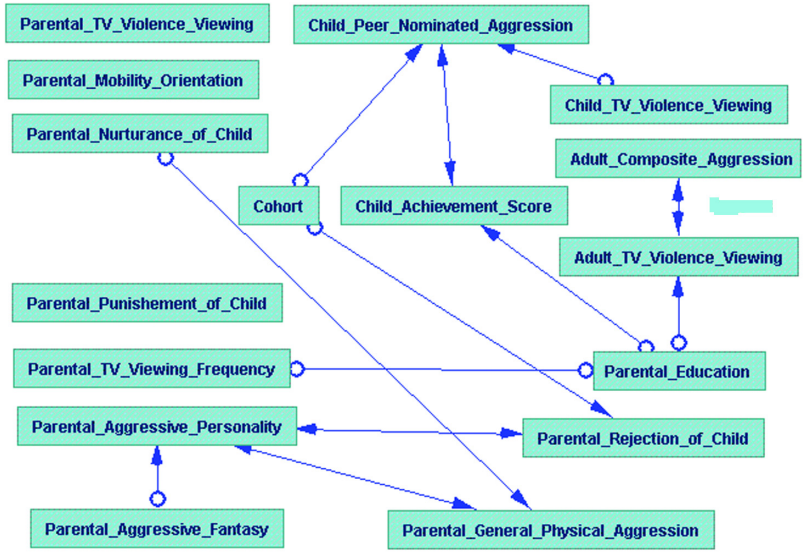


Note: Directed edges (arrows) represent a causal relation between the connected variables, with the arrow pointing to the effect; undirected edges represent the presence either of an unmeasured common cause or a direct causal relation between connected edges.

returned by FCI with $.1 \leq \alpha \leq .01$. That is, any causal model that explains the measured correlations and conditional independencies must not include a directed path, direct or mediated, from Child TV Violence Viewing to Adult Composite Aggression. Hence, if the measured associations and conditional independencies are representative, in females Child TV Violence Viewing is not a cause of Adult Aggression (see Figure 10).

In this respect, the patterns returned by GES and FCI searches are inconsistent with the models produced by Huesmann et al. (2003) for females. There is an explanation. Suppose two variables are connected by some path going through another variable *V*. If the edges on the path are directed into *V* (→ *V* ←), *V* is said to be a collider; all other nonterminal variables are mediating causes (→ *V* → or ← *V* ←) or common causes (← *V* →). If one conditions on all and only the colliders on a path between two variables, the variables will be conditionally associated. Note that in any model in the GES pattern, there will be a path from Child TV Violence Viewing to Adult Composite Aggression that includes Child Achievement Score as the only collider. Conditioning on this variable will therefore induce an association between Child TV Violence

Figure 10
FCI Output for Females From Huesmann Data



Note: Directed edges (arrows) represent a direct causal relation between the connected variables, with the arrow pointing to the effect; bidirected edges represent the presence of an unmeasured common cause; partially oriented edges (circle at one end and an arrowhead at the other) imply a direct causal relation or a shared unmeasured common cause; and unoriented (circle at both ends) edges imply that either the variables share an unmeasured common cause or one variable causes another but the algorithm cannot determine the direction of causation.

Viewing and Adult Composite Aggression. The second model Huesmann et al., parameterized (figure 6b, the model with a p-value above .2) assumed that Achievement was a common cause of Child TV Violence Viewing and Child Aggression. The possibility that Early Achievement could be a collider was not tested.

Conclusion: Interventions

Our analyses of the Huesmann data are consistent with what proponents of "convergence" argue, the case being stronger for males than for females. Given the assumption that the measured dependencies and independencies hold in the relevant populations, there likely, though not certainly, is a causal connection between exposure to televised violence and adult aggression. If there are no unmeasured common causes, the temporal constraints and the measured conditional independencies imply that exposure causes aggression, not vice versa. On the other hand, as far as we can

tell, the data themselves do not support an inference to the claim that there are no unmeasured common causes—this must, near as we can judge, be assumed if it is to be endorsed at all. Furthermore, for females, Huesmann data suggest that the relevant exposure is as an adult, not as a child.

Greater confidence in the causal structure relating television exposure to aggression, and in particular in the direction of included edges and the path coefficients, could be gained by the use of IV analysis using instruments for exposure, but finding valid instruments is not trivial. Aggressive search for natural experiments might produce more evidence similar to that of the Williams (1986) Canadian study. Furthermore, there are now a nontrivial number of designed field experiments that effect TV viewing (although in some cases, TV viewing was not the primary goal of the interventions). Such interventions could potentially be evaluated for effects on aggression. A different approach would focus on ruling out the possibility that aggression causes TV viewing by IV analysis using instruments for aggression. Including measurements of some cause of aggression that is clearly not a direct cause of exposure among either adults or children would provide greater confidence in the direction of the edges between exposure and aggression.

So, what policy conclusions can reasonably be drawn from the available studies, as matters stand? Suppose one concludes that watching violent TV in childhood does cause aggressive tendencies in some males, that watching violent TV in adulthood may cause aggressive tendencies in some females, and that those so affected are a sufficiently large proportion of the population for measurable correlations between those features. Can one predict the effect of abolishing violent TV for children in the general population? We think not really, for several reasons: (a) Because confounders cannot be eliminated, the size of the direct causal effects cannot be estimated—one can only test for sensitivity; (b) the actual dependencies may be nonlinear, so estimates of effect on a linear model would be wrong; and (c) the availability of substitute activities that may be related to the outcome has changed over time (e.g., video games).

Good estimates of the degree to which various levels of exposure increase risk of aggression, as measured by the frequency of aggressive acts of various kinds, are especially important for considerations of policy interventions. Such estimates, as noted above, are reliable only to the extent that one can properly condition on common causes. At the present pass, we are unwilling to endorse the assumption that there are no unmeasured common causes, in either the Johnson et al. (2002) or the Huesmann et al. (2003) studies.

Plato's solution was outright censoring of entertainment, but this is neither socially desirable nor politically feasible. State-enforced censorship has large social costs, whether or not it is effective. Moreover, the evidence available to us, at any rate, does not decide the questions whether exposure to media violence has an unequivocal causal effect on adult aggression in both genders and whether it is exposure as a child or as an adult that matters. For both reasons, less invasive interventions are preferable.

There are two obvious alternatives. The first is to encourage those with influence (e.g., parents, physicians, and teachers) to recommend against exposure and to reinforce such recommendations with public awareness campaigns. Whether such interventions would be effective, it might be reasonably supposed that they are likely to be more effective in reducing childhood than adult exposure.

A different alternative is to provide, as a matter of public policy, competing entertainment in the form of alternative programming options and nonviewing recreational options. This last option is especially appealing for three reasons. First, there is evidence suggesting that TV viewing has numerous other undesirable social consequences, especially for children, such as sedentarism, poor nutrition, and obesity (Gortmaker et al., 1996; Kahn, Ramsey, & Brownson, 2002; Marshall, Biddle, Gorely, Cameron, & Murdey, 2004; Shannon, Peacock, & Brown, 1991). The feasibility and effectiveness of interventions to reduce TV watching—usually with the goal of reducing obesity—has been demonstrated in several studies based in schools, preschools, and primary care settings (Dennison, Russo, Burdick, & Jenkins, 2004; Ford, McDonald, Owens, & Robinson, 2002; Gortmaker, Peterson, & Wiecha, 1999; Robinson, 1999). Second, these interventions arguably have low cost and positive benefits, both for individuals and for society, whether exposure to televised violence is a cause of aggression. And third, if exposure to televised violence does cause aggression, the social benefits of such interventions may be very large.

Anachronism or not, Plato may have been right about television, and about a lot else in popular culture that appears to be destructive. But Plato had no argument. Showing as much scientifically is a Herculean task, and we tip a hat to those who would argue for public policy on the best scientific ground they can muster rather than on the basis of religion or philosophy.

Notes

1. There are any number of particular methods by which one may statistically control for variables, all having in common a basic aim: to identify whether two variables are statistical independent given (conditional on) the value of some third variable or set of variables. We use such terms (e.g., controlling for, conditioning on, conditional on, adjusting for, and stratifying by) to indicate the general procedure rather than any particular statistical method for doing this. When particular statistical methods are at issue, we call them by name.

2. These fundamental notions are given various interpretations by different researchers. For example, one might say that X is a direct cause of Y with respect to V if and only if after some hypothetical interventions that fix values of all variables in V except for X and Y, Y covaries with randomized X. Many researchers view f as a measure of (someone's) degree of belief, whereas others argue that the conditional distributions implied by f by conditioning on (statistically controlling for) the variables in V that have no direct causes are approximations of the large sample distributions that would result from randomizing those exogenous variables. These differences are not important for our arguments in the remainder of the article, and we do not exclusively endorse any of these interpretations.

3. Since this paper was written, search procedures that find unique linear, graphical causal models with non-Gaussian distributions have appeared; the methods exploit higher moments and do not require Faithfulness. Since the data available to us are reported only as second moments–correlations–we have

been unable to make revisions that take advantage of these procedures (see Shimizu et al. (2006)). In our simulations, the output of the search procedures we describe here using only second moments produce results consistent the new methods for a variety of non-Gaussian distributions when Faithfulness holds.)

4. We are not here distinguishing between confounding and mixed populations. There are contexts in which the distinction is important; this is not one of them.

5. This is possible: The phenomenon is known as *hysteresis* and occurs with some frequency in, for example, population biology and latency models in epidemiology.

6. A variable *V* whose value is stable across a temporal span *t* to *t'* may be a common cause of some further variable *X* measured at *t'*, and changes in a third variable *Y* over the span *t'* to *t"*. This will occur if the value of *V* at *t* causes the value of *X* at *t'*, and also initiates a process beginning at *t'* and ending before or at *t"*, which process causes a change in the value of *Y*.

7. We thank Rowell Huesmann and his collaborators, Jessica Moise-Titus, Cheryl-Lynn Podolski, and Leonard Eron, for making the matrices available. We should have liked to do the same for the Johnson, Cohen, Smailes, Kasen, and Brook (2002) study, but our request for the relevant data was denied.

8. Strictly, the output of FCI is a Partial Ancestral Graph (PAG). If there is an edge from *X* directed into *Y*, then no matter the mark at the *X* end, *Y* is not an ancestor of *X* in the causal graph—not a cause of *X*—and there is an association between *X* and *Y* that cannot be removed by conditioning on any subset of the observed variables. Not all edges in a PAG need denote edges in the true graph of causal relations.

9. The edge is included in patterns returned at $.1 \geq \alpha > .05$, with no other edges into or out of Adult Composite Aggression. It is lost with $\alpha < .045$, so that Adult Composite Aggression is unconnected in the pattern; at $\alpha < .04$, Child TV Violence Viewing also becomes isolated in the returned pattern.

10. For $\alpha = .1$ and $\alpha \leq .2$, the edge between Adult Composite Aggression and Adult TV Violence Viewing is unoriented, permitting a common cause, or a direct causal relation in either direction; at $\alpha = .25$, the edge is partially oriented into Adult TV Violence Viewing.

References

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69-85.

Dennison, B. A., Russo, T. J., Burdick, P. A., & Jenkins, P. L. (2004). An intervention to reduce television viewing by preschool children. *Archives of Pediatrics & Adolescent Medicine*, *158*(2), 170-176.

Ford, B. S., McDonald, T. E., Owens, A. S., & Robinson, T. N. (2002). Primary care interventions to reduce television viewing in African-American children. *American Journal of Preventative Medicine*, *22*(2), 106-109.

Geen, R. G., & Donnerstein, E. (1998). *Human aggression: Theories, research, and implications for social policy*. San Diego, CA: Academic Press.

Gortmaker, S. L., Must, A., Sobol, A. M., Peterson, K., Colditz, G. A., & Dietz, W. H. (1996). Television viewing as a cause of increasing obesity among children in the United States, 1986–1990. *Archives of Pediatrics & Adolescent Medicine*, *150*(4), 356-362.

Gortmaker, S. L., Peterson, K., & Wiecha, J. (1999). Reducing obesity via a school-based interdisciplinary intervention among youth—Planet health. *Archives of Pediatrics & Adolescent Medicine*, *153*(4), 409-418.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, *29*(4), 722-729.

Huesmann, L. R., Moise-Titus, J., Podolski, C.-L., & Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology*, *39*(2), 201-221.

Johnson, J. G., Cohen, P., Smailes, E. M., Kasen, S., & Brook, J. S. (2002). Television viewing and aggressive behavior during adolescence and childhood. *Science*, *295*(5564), 2468-2471.

Kahn, E. B., Ramsey, L. T., & Brownson, R. C. (2002). The effectiveness of interventions to increase physical activity—A systematic review. *American Journal of Preventive Medicine*, *22*(4), 73-108.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, *48,* 371-410.

Luketsch, H. (1989). Video violence and aggression. *German Journal of Psychology*, *13,* 293-300.

Marshall, S. J., Biddle, S. J. H., Gorely, T., Cameron, N., & Murdey, I. (2004). Relationships between media use, body fatness and physical activity in children and youth: A meta-analysis. *International Journal of Obesity Related Metabolic Disorders*, *28*(10), 1238-1246.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causation*. Cambridge, UK: Cambridge University Press.

Pecora, N., Murray, J. P., & Wartella, E. (in press). *Children and television: 50 years of research*. Mahwah, NJ: Lawrence Erlbaum.

Robins, J., Scheines, R., Spirtes, P., & Wasserman, L. (2003). Uniform consistency in causal inference. *Biometrika*, *90*, 491-515.

Robinson, T. N. (1999). Reducing children's television viewing to prevent obesity—A randomized controlled trial. *Journal of the American Medical Association*, *282*(16), 1561-1567.

Rubenstein, E. A. (1982). Introductory comments. In D. Pearl, L. Bouthilet, & J. Lazar (Eds.), *Television and behavior: Ten years of scientific progress and implications for the eighties* (p. 104). Rockville, MD: National Institute of Mental Health.

Shannon, B., Peacock, J., & Brown, M. J. (1991). Body fatness, television viewing and calorie-intake of a sample of Pennsylvania 6th grade children. *Journal of Nutrition Education*, *23*(6), 262-268.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation prediction and search*. New York: Springer-Verlag.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation prediction and search* (2nd ed.). Cambridge, MA: MIT Press.

Stacy, L., Smith, S. L., & Donnerstein, E. (1998). Harmful effects of exposure to media violence: Learning of aggression, emotional desensitization, and fear. In R. G. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 168-202). San Diego, CA: Academic Press.

Williams, T. (1986). *The impact of television: A natural experiment in three communities*. Orlando, FL: Academic Press.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Yanez, N. D., Kronmal, R. A., & Shemanski, L. R. (1998). The effects of measurement error in response variables and tests of association of explanatory variables in change models. *Statistics in Medicine*, *17*(22), 2597-2606.

Yanez, N. D., Kronmal, R. A., Shemanski, L. R., & Psaty, B. M. (2002). A regression model for longitudinal change in the presence of measurement error. *Annals of Epidemiology*, *12*(1), 34-38.

Zhang, J., & Spirtes, P. (2003). *Strong faithfulness and uniform consistency in causal inference*. Proceedings of the 2003 Conference on Uncertainty in Artificial Intelligence.

Bruce Glymour is an associate professor of philosophy at Kansas State University. He works on explanation and model selection in population biology and selected social sciences.

Clark Glymour is an alumni university professor at Carnegie Mellon University and senior research scientist at the Florida Institute for Human and Machine Cognition.

Maria Glymour is currently a Robert Wood Johnson Health & Society Scholar at Columbia University and an assistant professor in the Department of Society, Human Development, and Health at the Harvard School of Public Health. Her research focuses on social determinants of health.