**How PC Can Discover “Constitution”**

Clark Glymour[[1]](#footnote-1), Joseph Ramsey, Shuyan Wang

Carnegie Mellon University

Abstract

In an exchange with Gebharter, Casini and Baumgartner (B and C) conclude that the PC algorithm cannot reliably discover the “constitution” of materials or systems and has not been used to do so. We address their arguments, show sufficient conditions for PC to discover composition from data framed as Gebharter proposes, describe how in scientific practice using frameworks for data different from Gebharter’s PC has discovered constitution in two domains and two senses, and explain why PC is more reliable than standard methods used for causal or constitutional inference with non-experimental data. We address B and C’s examples in Appendix 2.

1. From Mental Causation to the Incapacities of a Search Algorithm

A recent exchange between Gebharter (2017,a,b), on one side, and Baumgartner and Casini (2023) (hereafter: B and C) on the other, illustrates one of the charms of philosophy: start with one problem and who knows where you will end up?

Gebharter started with the problem of mental causation: if the mental supervenes on the physical—same physics then same mental--can mental properties (or events or processes) cause anything? In the course of addressing the question, Gebharter argued that mind and body relations have the same probabilistic structure as directed acyclic graphical models used in causal analysis. Gebharter proposed that a search algorithm that exploits that structure, PC (Spirtes and Glymour, 1991), might be applied to attempt to reveal constitution, noting a difficulty with deterministic relations that B and C subsequently emphasized.

Challenging Gebharter’s argument led B and C to the following thesis:

A. *“*On the basis of two benchmarking experiments, we show that *even under discovery circumstances that—apart from the presence of determinism—are maximally favourable to the performance of PC, the algorithm’s capacity to correctly recover constitutive relations is*

*not high enough to counterbalance the severe risk of false positives.”*

And this claim:

B. “*While PC is one of the most frequently discussed causal discovery tools, it has played no role so far in constitutive discovery.” [[2]](#footnote-2)*

Why dispute over a particular algorithm turning on liabilities that were recognized thirty years ago and compared to which more accurate and more general algorithms have long been available? PC became the focus of the exchange when Gebharter proposed its use because it is an old and familiar algorithm, but PC is the ancestor of a number of search algorithms—including algorithms that use quasi-Bayesian scores rather than hypothesis tests--that like PC exploit selective conditional probability relations among measured variables, assumed to reflect causal relationships (or their absence). The essentials of Gebharter’s proposal would endure if some such related algorithm other than PC were able to discover constitution. B and C’s complaints are really with this whole class of statistical search procedures because, although their simulations are almost exclusively with the PC algorithm, their arguments, if valid, would impugn the entire class.

Our aim in what follows is to dispute A and B above.

There is an ambiguity. By the “constitution” of something can be meant what it is composed of, what makes it up, its constituents or components. But something else might be meant by “constitution”: what do the components do and how are they related? The second sense is about causality. B and C’s examples range over both senses.

Gebharter frames the data for discovering constitution as a distribution of cases, in some of which the composite and components are present, and in some of which only the components but not the composite are present. Within Gebharter’s framing of how data bear on composition, we provide sufficient conditions for PC to discover what a composite is composed of. Gebharter’s framing of the kinds of data for discovering composition is, however, special. There are other data framings, and PC has been used with one of them to discover what materials—in this case rocks and soil--are composed of. In still another data framing, PC has been used successfully to discover what components—in this case, genes—do.

Claim B, italicized above, is demonstrably false in both senses of “constitution.” Claim A is less precise and depends on what “high enough to counterbalance the severe risk of false positives” means and how that would be judged. If, as in many empirical studies, one expects that measured variables have unmeasured causes that contribute to variation from case to case in the values of the measured variables, then the relations among the measured variables will not be deterministic. B and C’s major criticism—that PC will not work when deterministic relations are present--will not apply. Determinism will hold if one variable is specified to be a function of one or more other variables, for example if X is a variable and Y is specified to be 2X or Z is specified to be the sum of X and Y. Such relations are often deliberate, made by definition, and where that is known, simple programming can instruct PC not to condition on one or more entangled variables when estimating connections of the others.[[3]](#footnote-3) Alternatively, “high enough” might be judged by comparison with methods commonly used in the sciences for discovering causal or constitutional relations from observational data. The most common methods are varieties of regression, and we will make that comparison. Finally, we note what we regard as the most serious limitations of PC and some of its algorithmic friends: unobserved common causes, mixed data, sample selection bias and mismeasurement.

1. The Arguments, in Nutshells

The B and C argument, illustrated with multiple simulations, is this:

1. Lots of deterministic relations are mathematically possible;

2. constituitive relations are deterministic;

3. if X determines Y and Y causes Z, PC will not find a Y – Z connection, which will often lead to other errors;

4. if X, Y, Z, etc. are related by a deterministic equation, e.g., any two of them determine the third, , then PC will miss connections with and among those variables ;

5. *therefore, the algorithm’s capacity to correctly recover constitutive relations is*

*not high enough to counterbalance the severe risk of false positives.[[4]](#footnote-4)*

Determinism of Y by a set **Z** of variables we understand to mean that for each assignment of values to members of **Z**  that occurs in the data, Y has the same value for all members of **Z** occurring in the data and sharing that assignment. False positives we understand in two senses: including an edge representing a non-existent causal connection and orienting a connection in the wrong direction. We note that if PC adds or misses an edge, the program may mistake the orientations of other edges.

The first of these premises is of course true, and the third is true and was noted by the authors of the PC algorithm two years after the algorithm appeared (Spirtes, et al., 1993). The fourth is true, but, as we will argue, not obviously related to what is a constituent of what. The second premise is false, and the conclusion is false. Here are the counter-arguments[[5]](#footnote-5):

1. Suppose the composite and the constituents are measured and the data are as Gebhardt and B and C frame the problem of discovering composition. Suppose there are two or more constituents and the joint presence of all constituents determines the presence of the composite and the absence of any constituent determines the absence of the composite. Then, if for every constituent there is another constituent independent of it (in probability, or as correctly estimated from frequencies in a sample), or independent of it conditional on any other set of measured variables (e.g., common causes or common constituents), and there are no unmeasured common causes, and there is no sample selection bias, PC identifies the composition relations.
2. Composition is semi-deterministic, not deterministic, and semi-deterministic relations do not have the problem that PC cannot find a Y – Z connection if X determines Y.
3. B and C overlook techniques outside of Gebhardt’s framing for estimating the composition of materials and systems in which PC has shown to be reliable and informative in real scientific problems.

All the reader needs to know about the PC algorithm for the following discussion is this: the procedure seeks a set of acyclic directed graphs—sometimes called DAGs--whose vertices are random variables. The set sought is all DAGs on the measured variables entailing the same conditional independence relations—a Markov equivalence class having the true DAG as a member. The algorithm starts with a complete undirected graph and sequentially removes each edge X – Y for which it finds a minimal set of variables with edges connected to X or to Y for which X and Y are independent conditional on all values of the variables in the set. When the first stage procedure terminates, a second stage directs each pair of edges X - Y – Z for which there is no X -Z edge as X -> Y <- Z if in the first stage Y was not a member of the set of variables conditioned on in removing the X – Z edge. Further consequences of such orientations are developed.

The algorithm is provably correct[[6]](#footnote-6) (given true information about conditional independence) under a number of jointly sufficient—but not necessary--assumptions, one of which is that there be no deterministic constraints on the measured values of the variables. Whether the edge relations signify causes or composition or something else is left to the user.

1. PC Recovers Deterministic Composition with Two or More Independent Constituents

In Gebharter’s framing, a constitution problem consists of data containing cases, each case specifying a value for the composite and a value for each candidate constituent. Taking the variables to range over two possible values, Present and Absent, and taking the actual constituents of Y to be X1, X2, X3, X4 and X5, mutually independent in probability, with a sample of 1,000 cases and the default search settings[[7]](#footnote-7), PC returns Figure 1.

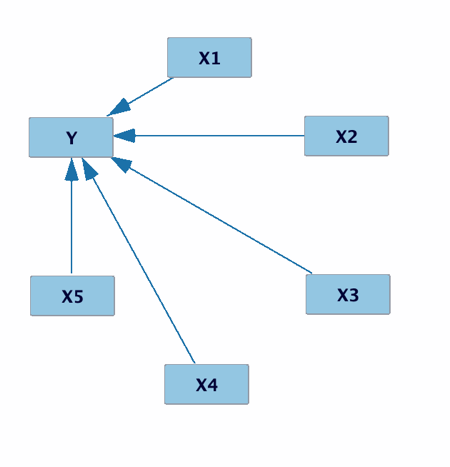


Figure 1. A PC result with binary variables when P(Y) = 0 if and only if at least one of the X variables is 0 and P(Y) = 1 when all X variables are 1. N = 1,000.

Proofs for general cases are given in the appendix.

.

1. Composition is semi-deterministic, not deterministic

If X, Y are part of the composition of Z, then Pr(X, Y both present |Z = present ) = 1, or equivalently, Pr(Z = present | X = absent or Y = absent) = 0. We will call such a relation *semi-deterministic* when in addition: 1 ≠ Pr(X, Y both present |Z = absent ) ≠ 0 . Containment and composition are semi-deterministic: whatever is sulfuric acid must contain sulfur and oxygen and hydrogen, but none of these constituents alone is sulfuric acid. And so on.

Except for special cases, if X only semi-determines Y, then conditional on X, Y and Z remain dependent as v*ariables* ifthey were dependent unconditionally—for some value of X, Y and Z are dependent. Given the true independencies from data as in Gebharter’s framing, PC will find the Y - Z connection.

There are special cases. Suppose the structure of Figure 2 and that binary X1 semi-determines Y and W, both binary, but for different values of X1. Pr(Y = 1 | X1 = 1) = 0 = Pr(W = 1 |X1 = 0). Then conditional on any value of X1, one of Y, Z will be constant, and PC will not find the Y – W connection.

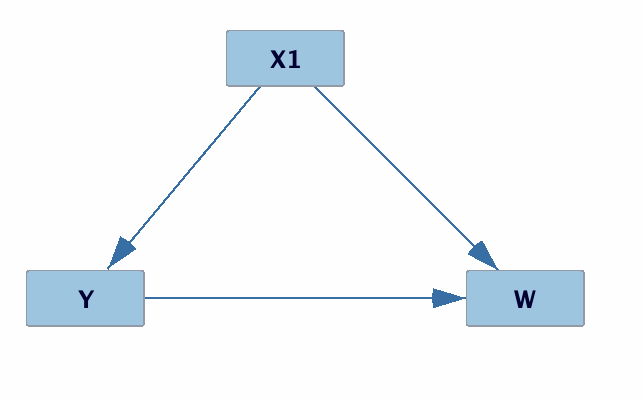


Figure 2. A semi-determinism problem for PC

1. Techniques for estimating the constitution of materials

5.1 Gebharter cites PC as an imagined way to analyze mental causation, and B and C adopt his data framing. In Gebharter’s framing, application of PC to estimate composition would require data in which a description of the data itself specifies that

* there are cases in which both the composite and the constituents are present
* as well as cases in which one or more of the constituents is present and the composite absent
* and no cases in which the constituent is present and one or more of the constituents absent.

Such a data set would scarcely warrant a statistical search of any kind. The conclusion would follow for the sample almost from the very description of the data. That sort of data does occur, notably in studies of chemical composition, where the real work is creating, identifying and measuring the materials. The 18th and 19th centuries saw the identification of compounds by their properties, interactions, decomposition and synthesis, and of elements by their non-decomposability and as reactants in synthesizing compounds, and by other features. Basically, an element was declared when it was found that no process separated a sample of a substance into two or more others the sum of whose weights equaled the weight of consumed reactants. Inference was without formal probability. Lavoisier’s *Treatise* contains many claims that one thing or another is probable, but has no probability calculations, although, Laplace, his collaborator, wrote the foundational work on probability and used it in a hypothesis test about the planes of the orbits of the planets. Aspects of the structure of compounds were eventually found by Cannizzaro’s ingenious method of estimating atomic weights. Notably, none of this work involved statistical inferences from frequencies. Statistics had no place in Cannizzaro’s famous paper that helped Mendeleev to his still more famous periodic table.

But chemical composition is also one of the domains in which the PC algorithm has proved itself.

5.2 Spectroscopy and a Different Data Framing for Discovering Composition

After Fraunhofer’s invention of the spectroscope around 1814, spectroscopic methods began to be applied to determine composition in gases, the sun and other targets. A variety of spectroscopic techniques are now commonly used to estimate chemical composition for planetary atmospheres and stars, for land cover estimated from satellite data, and for chemical or biological composition for materials rather closer to the spectroscope. There are many techniques, for example absorption spectroscopy, reflectance spectroscopy, Raman spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, and more. Characteristic of spectroscopy is an array of signals--intensities at various frequencies in light spectroscopy. Identification of composition is done by identifying particular signals (or their absence) at frequencies characteristic of a substance or in more complex cases by a pattern of relative intensities of signals at some interval of frequencies. Some techniques, such as NMR, estimate the relative abundance of chemical species. Libraries of reflectance and Raman spectra of known materials are kept for comparison with objects to be analyzed.

Thanks chiefly to the work of the late Phoebe Hauff of the United States Geological Survey, by the middle of the 20th century reflectance spectroscopy became a common method for estimating the mineral composition of rocks and soils. Reflectance spectroscopy measures the intensities of light of various frequencies reflected from a surface, often where possible standardized against the spectrum of white light near the same source. Deciphering the spectra of natural materials is challenging. Distinct minerals appear together in natural formations and share reflectance frequencies. A standard reference lists more than 4000 minerals and more than 100 carbonates. Reflectance spectra vary with illumination and atmospheric conditions. Reflectance spectra from natural objects are therefore a complicated effect of the composition of a surface and the environment. Mineral spectroscopy thus invites the use of statistical methods for inferring constitution. Traditionally, standard statistical methods such as multiple linear regression have been used, and more recently there have been proposals for using various machine learning methods. This is a context in which it would be appropriate to test PC’s “capacity to recover constitution.” We have evidence.

Ramsey, et al., (2002), conducted a series of studies for the National Aeronautics and Space Administration to estimate mineral composition from reflectance spectra. Identification of carbonates was of particular interest because they are often deposited from water, and water was considered essential for past life on a planet. The project had three parts. One was to identify carbonate minerals from spectra taken *in situ* in the Mojave desert of southern California. A second was to distinguish carbonate from non-carbonate objects planted for the purpose in a field at NASA’s Ames Research Center. The third was a comparison of the informativeness and accuracy of a human expert and the PC algorithm for identifying carbonates from the spectra of an extensive collection of materials of known composition maintained at the Johns Hopkins University. In all of these cases the task was to determine some of the composition of novel objects.

A modification of the PC algorithm[[8]](#footnote-8) used as data the spectrum of a target object together with spectra from a library of known pure minerals at the Jet Propulsion Laboratory, with each spectrum labeled by its mineralogical source. PC was a recourse after initial experiments with neural nets performed poorly on minerals outside the training set. The PC search sought to estimate whether the spectrum of the novel object could be attributed to any of the carbonate minerals in the JPL library. That is, graphical model edges were sought between the label of the target object and the labels of the library minerals, using the intensities of spectra at multiple frequencies as data. The data table was quite different from Gebhardt’s formulation. It had rows labeled by spectral frequencies and columns labeled by known minerals and one column labeled as the target. The cell entries were the intensities of the spectrum for the row frequency and column mineral from a reference library, but with the cell entries for the target object taken from experimental or field measurements of its spectrum. Some prior knowledge specifying the interval of frequencies most characteristic of carbonates was used to restrict the data, which was also pre-processed, smoothing and removing spikes (the later due to atmospheric water), in ways blind to composition of the target object.

Some of the samples from the Mojave were analyzed chemically, and all samples were assessed for carbonate content by experts *in situ* from appearance and spectral properties. The algorithm results almost perfectly matched the expert assessments of samples for the presence or absence of two specific carbonates, calcite or dolomite, and on the presence or absence of carbonates of any kind agreed in 15 of 20 samples. PC also agreed with the subset of samples chemically tested for carbonate content.

A second study took reflectance spectra from objects of known composition planted at NASA Ames Research Center. With a few scattered point errors outside the objects, modified PC correctly identified carbonate and non-carbonate regions. Multiple regression was essentially useless, scattering “C” all over the field of view.

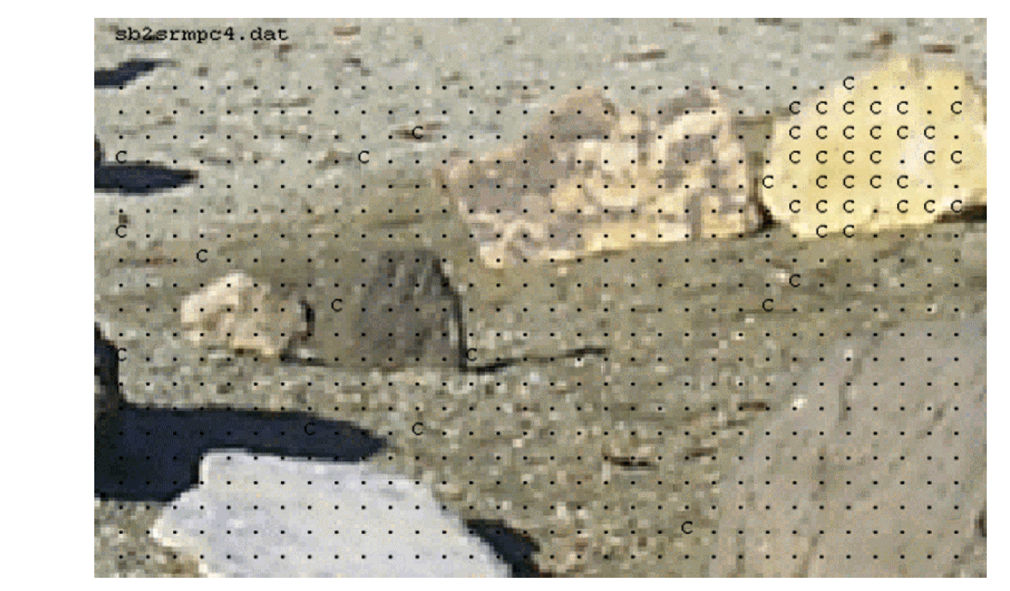


Figure 3. PC estimation of carbonate surfaces marked by “C” from Ramsey, et al., 2002. Only the yellow rock at the top right is a carbonate.

A third study applied the modified PC algorithm to the Johns Hopkins Library of rock spectra in an informal competition with an expert spectroscopist, Ted Roush. Roush had access to the Johns Hopkins spectra but not to the mineral identities and could use any reference materials he wished and could take as much time as he wanted. In the event, he spent twelve hours on the task. PC took a few minutes on a work station of the era. The results were that for the 20+ calcite and dolomite minerals in the library, Roush was almost unerring, making only a single mistake. But he missed almost all of the rest of the carbonate minerals in the JHU library.[[9]](#footnote-9),[[10]](#footnote-10) PC, made more errors of omission with calcite and dolomite but correctly identified many more of the carbonate minerals with few false positives. Altogether, PC was a good apprentice spectroscopist, and better than standard statistical methods or neural nets--which tended, like Roush, to fail with mineral combinations of kinds it had not been trained on.

5.3 Discovering gene regulators.

Exploiting the ambiguity in “constitution,” some of B and C’s simulated examples are about discovering the relations among features of a system rather than discovering what is in the system. There are tests of PC in identifying such relations in natural systems, albeit where determinacies, if any, are unknown, and where, as in all empirical science, there is measurement error. Steckhoven, et al., (2011), applied PC within a subsampling and effect estimation ranking design they call CStar to observational data on yeast gene expression to identify regulatory genes. They then compared the results with established experimental determinations of gene regulation relations in the species. Their results for the PC method, CStar, compared with random guessing and commonly used statistical procedures (regression, LASSO, elastic net, simple correlation), and their method without effect estimation (IDA) are summarized in the receiver operating characteristic (ROC) curves shown in Figure 4 (their Figure 2).

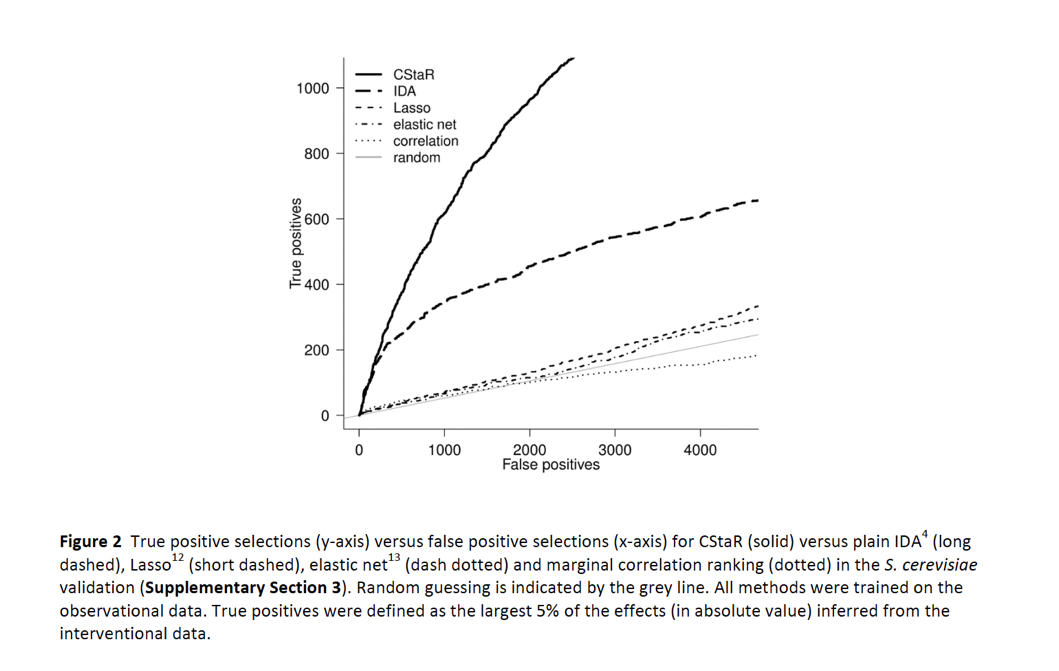


Figure 4. Steckhoven, et al., comparison of accuracy of a PC method (CStaR) combined with subsampling and effect estimation and other methods for identifying gene regulators in yeast.

Their experiment was repeated by an independent group with similar results. In a separate experiment Steckhoven, et al. applied the same method to search for genes that regulate flowering time for *A. thaliana* from among more than 20,000 candidate genes using as data for PC the gene expression profiles and flowering times for only 47 plants, discovering 5 known and 4 novel regulators among the top 5% of their ranking of genes, confirmed by subsequent gene knockout experiments. No false positives were obtained, but some plants with modified genomes did not survive.

1. Comparison with Regression

Another standard to which one might hold PC is comparison with standard methods for identifying composition or causality. Regression methods are widely used, and sometimes advocated for such purposes (Rawlings,1988)[[11]](#footnote-11). We have described their inferiority to PC in some particular empirical cases, but there are more general considerations. Whether linear or logistic or otherwise, regressions share the feature that the effect or composition of one variable for another is estimated by conditioning simultaneously on all other “explanatory” variables. PC, as noted, conditions only on selected subsets; other subsequent “greedy” algorithms also consider only selected subsets of other variables.

Regression cannot be used to estimate causality or composition (or anything) when then number of prediction variables exceeds the number of data points. PC and related greedy algorithms can in most cases. The Fast Greedy Equivalence search, for example, an algorithm that exploits conditional independence relations using a quasi-Bayesian score, has recovered more than 90% of a million variable simulated structure with only 10% false positives in a problem with one thousand cases in the data.[[12]](#footnote-12) (Ramsey, et al. 2017).

Regression fails when a proper subset of the predictors suffices to account for almost all of the variance of the target variable. PC does not if it does not select such a set for conditioning. For related reasons, PC reduces (but does not eliminate) problems of multicollinearity that arise when a multitude of prediction variables determine, or almost determine, the value of another predictor.

When a single predictor variable, X, shares an unmeasured common cause with an outcome variable Y, other predictors that cause X or share an unmeasured common cause with X, but do not cause Y, regression will estimate that these other predictors directly influence Y. If the associations are strong but not deterministic, PC will not.

1. The Real Problems

PC requires conditional independence tests. Appropriate tests depend on the joint distribution of the variables. A general test, appropriate for a wide range of distributions, is available, but its time requirements increase with the number of data points and therefore it can be very slow.

PC assumes there are no unobserved common causes, and in real data, especially observational data, that is often a dubious assumption. PC is defeated by certain kinds of measurement errors, in which cases it does find false positives.

Data may be from a mixture of distinct causal structures with distinct probability distributions. At least for Gaussian data, there is a published solution (Zhang and Glymour, 2020).

Finally, the sample may be biased, for example if membership in the sample is dependent on values of some of the measured variables (Zhang, et al., 2016).

1. Back to Gebharter

We intend only to respond to many of B and C’s criticism of PC, not to support Gebharter’s specific proposal for discovering the constitution of the mind. It has previously been proposed that mental processes accord with Bayes net or dynamical Bayes net representations (Glymour, 2001; Gopnik, et al., 2003), but even were that true it would not warrant using methods such as PC for discovering such structures. One reason is that psychological data do not accord with Gebharter’s data framework. Except in introspectionist methods that are now widely distained in psychology, data do not provide values of mental states, only of human behavior in experimental or observational circumstances or of physical effects of brain physiology in experimental situations, as for example in functional magnetic resonance imaging. There are, however recent algorithmic methods for inferring unmeasured causes of measured variables with specific distribution families (Zie, et al., 2022; Kong, et al., 2023).

1. Conclusion

An eminent philosopher once told one of us that philosophy is destructive testing.[[13]](#footnote-13) B and C undertake such a service for PC and kindred algorithms. Unfortunately, many of the claims and complaints B and C lodge against the capacity of the PC algorithm to recover composition are mistaken. PC cannot discover “constitution” relations in data framed as Gebhardt imagines, but its limitations for inference to composition in that framework are essentially the same as those for causal inference and largely unrelated to the criticisms of B and C. B and C wrongly claim that PC has never been successfully used to discover constitution. Moreover, Gebhardter’s framing is special. Quite different data acquisition and representation schemes are possible, are used in practice, and PC succeeds in them—indeed succeeds better than more commonly used statistical procedures. When a new algorithm more accurate or more informative than PC appeared, Peter Spirtes, the P in PC, once remarked that everything beats PC. Not quite. [[14]](#footnote-14)

References

Andrews, B., Ramsey, J., Sanchez-Romero, R., Camchong, J., & Kummerfeld, E. (2023). Fast Scalable and Accurate Discovery of DAGs Using the Best Order Score Search and Grow-Shrink Trees. In Proceedings of the Neural Information Processing Systems (NeurIPS) 2023.

Baumgartner, M. and C. Cassini, (2023) The PC Algorithm and the Inference to Constitution. *British Journal for the Philosophy of Science*, 405-429l.

Gebharter, A. (2017a). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies,* 174, 2645-2666.

Gebharter, A. [2017b]. Causal Exclusion and Causal Bayes Nets. *Philosophy and Phenomenological Research*, 95, pp. 353–75.

Glymour, C. N. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. MIT press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological review*, *111*(1),

Kong, Lingjing, Biwei Huang, Feng Xie, Eric Xing, Yuejie Chi, Kun Zhang, [“Identification of Nonlinear Latent Hierarchical Models,”](https://neurips.cc/virtual/2023/poster/70653) accepted to Conference on Neural Information Processing Systems (NeurIPS) 2023.

Ramsey, J., Glymour, M., Sanchez-Romero, R., & Glymour, C. (2017). A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International journal of data science and analytics*, *3*, 121-129.

Ramsey, J., Gazis, P., Roush, T., Spirtes, P., & Glymour, C. (2002). Automated remote sensing with near infrared reflectance spectra: Carbonate recognition. *Data mining and knowledge discovery*, *6*, 277-293.

Andrews, B., Ramsey, J., Sanchez-Romero, R., Camchong, J., & Kummerfeld, E. (2023). Fast Scalable and Accurate Discovery of DAGs Using the Best Order Score Search and Grow-Shrink Trees. In Proceedings of the Neural Information Processing Systems (NeurIPS) 2023.

Rawlings, J. O. (1988). Applied regression analysis: a research tool. Wadsworth & Brooks. *Pacific Grove, CA*.

Spirtes, P., & C. Glymour. (1991). An algorithm for fast recovery of sparse causal graphs, *Social Science Computer Review,* 9(1), 62-72.

Spirtes, P., C. Glymour and R. Scheines. (1993). Causation, Prediction and Search, Springer Lecture Notes in Statistics.

Stekhoven, D., Hennig, L., Sveinbjörnsson, G., Moraes, I., Maathuis, M., & Bühlmann, P. (2011). Causal stability ranking. *Nature Preceedings*, 1-1.

Zhang, K., & Glymour, M. R. (2020). Unmixing for causal inference: Thoughts on mccaffrey and danks. *The British Journal for the Philosophy of Science*.

Xie, F., Huang, B., Chen, Z., He, Y., Geng, Z., & Zhang, K. (2022, June). Identification of linear non-gaussian latent hierarchical structure. In *International Conference on Machine Learning* (pp. 24370-24387). PMLR.

Appendix 1

PC discovers composition when for each component there is another component of which it conditionally independent and also when some of the variables are components of more than one composite.

A simple proof when the components are assumed to have a common cause Z.

Let C and X1,…,Xn, n ≥ 2 be jointly distributed as

P(C, X1,…,Xn) = P(C | X1,…,Xn) P(X1 | Z)…P(Xn | Z)P(Z)

P(C = 1 |X1,…,Xn = 1) = 1 and 0 otherwise

For all values of Xi, Z, 0 ≠ P(Xi), P(Z) ≠ 1

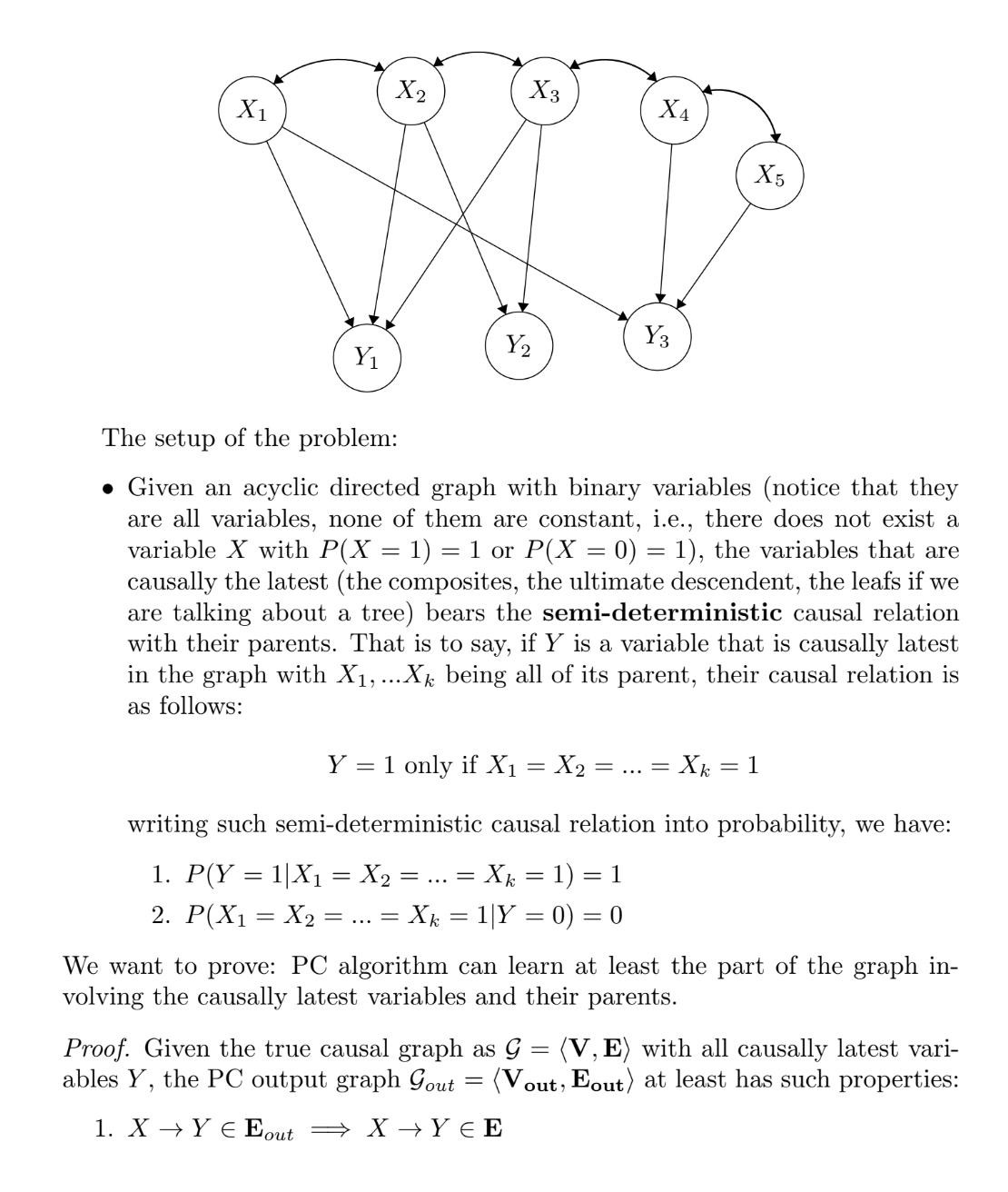
Then, given conditional independencies in accord with 1, PC finds a graph with an edge directed from each Xi to C and no other edges among {X1,…,Xn,C}.

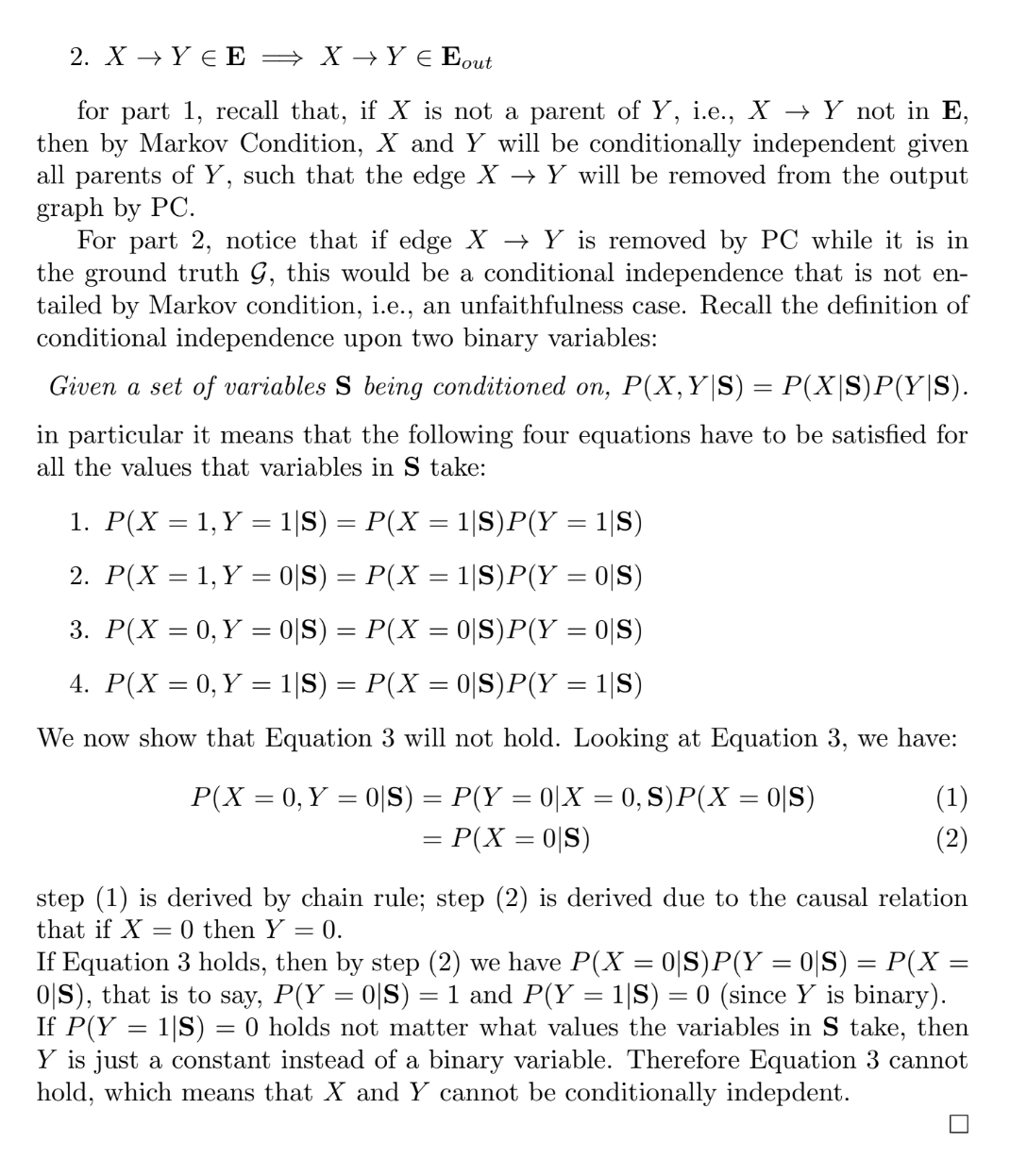
Proof: All Xi – Xj edges are removed from the initial PC undirected graph, because each Xi is independent of each Xj unconditionally or conditional on Z. Undirected edges between Xi and C remain because Xi and C are not independent conditional on any subset of variables not containing Xi or C. Xi – C – Xj is therefore in the final result of the first stage of PC, and C was not conditioned on in removing the Xi -Xj edge. Therefore, PC directs the edges as Xi -> C <- Xj.

The essential idea is that PC orients “uncovered” colliders: Xi -> Y <- Xj.

If components jointly determine a composite then whether PC finds effects of the composite is chancy and sensitive to parameter settings in the data generating model.

A more detailed proof when the components have separate associations.





Finally, we note that by conditioning on the *X* variables the argument can be iterated through a third layer of which *Y* variables are components.

.

Appendix 2: Comments on B and C Examples

In order not to interrupt the flow of argument of the paper, we consider here several examples B and E offer that pose difficulties for PC. We note that the presence of deterministic relations is readily detected in many cases, for example with linear relations the correlation matrix is not positive definite and not invertible.

Example 1

“Casting a vote, W =1, can be constituted by a raise of either the left hand, L = 1, or the right hand, R = 1 (but raising both hands is invalid…This system of binary variables features not only bottom-up determination but also top-down determination: any of the four possible value conﬁgurations of {W, L} and of {W, R} determine the value of R and L, respectively. Hence, not only the phenomenon of voting but also the hand raisings are screened off from all variables outside of that system. But hand raisings, for example, have causes in the motor cortex and effects in air displacement, meaning that outside variables can de facto causally interact with R and L. That these outside variables can be screened off from R and L in mixed variable sets comprising complete sets of constituents, therefore, violates CFC.”

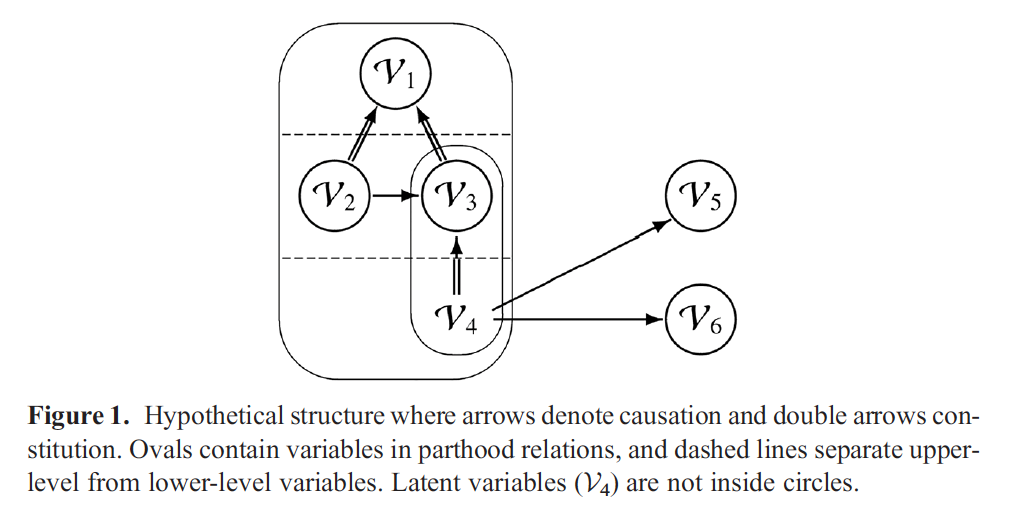
There is a lesson in the example, but not the one intended. The lesson is that correctly applying “BN axioms” or any other method depends on forming variables appropriately. “Screening off” is a relation among variables. Variables have mutually exclusive values and more than one value. It is misleading terminology to say that a value of a variable “screens off” other values of that variable. If genetics causes height, a height of 5 feet does not screen off genetics from height of 6 feet. In the authors’ example, Voting is presumably a variable taking values 0 or 1; Handraising is a variable taking values in {L, R}. Handraising screens off its causes from Voting, but Voting does not screen off Handraising from Outside causes. And if everyone votes, voting is constant and independent of everything.

One can always divide a variable with a range, R, into two or more logically connected variables with mutually exclusive subsets of R as ranges and then claim “screening off relations.” That is what B and C have done. What failed for their Bayes net analysis is how B and C divided up the world.

Continuous variables that are thought to be potential non-linear causes of an outcome variable are sometimes fragmented into separate variables in order to use a conditional Gaussian or multinomial distribution, and outcome variables have sometimes been separated into binary or multiple categorical variables in order to use logistic regression. Any wise data analyst takes regard for the dependencies and conditional independencies thus induced.[[15]](#footnote-15)

Example 2

B and C give an example, shown in their Figure 1,



Assuming the double headed arrows are deterministic, B and C show by simulation that PC cannot recover the structure from simulated values of the variables. The example is puzzling. If V4 uniquely determines the value of V3, how can V1 be a cause of V3? That aside, even if the double headed arrows are semi-deterministic, B and C are right for some combinations of parameter values, but in part for an entirely irrelevant reason. In Figure 1, V4 is an unmeasured cause of V5 and V6, and, quite generally, PC cannot discover causal relations between unmeasured and measured variables. Determinism and semi-determinism have nothing to do with that. It is worth exploring how PC behaves without various disabling conditions for Figure 1.

If the double headed arrows represent nearly semi-deterministic relations and V4 is measured but, as there would be with measurement error, there is the slightest variance in V3 and V1 conditional on the absence of their respective constituents, PC can recover most of the structure provided the other associations are statistically detectable.

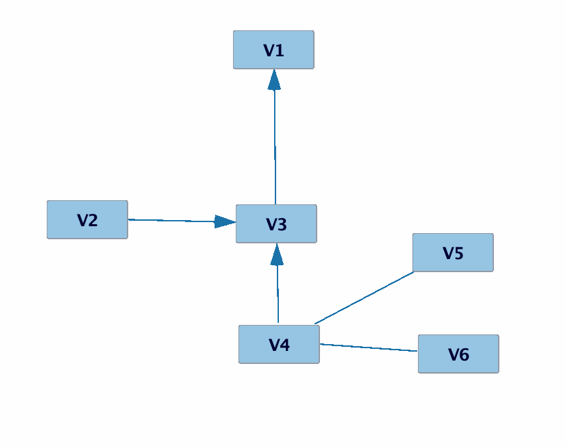


Figure 2. PC result with semi-determinism, measured V4 and 1% variance of V1 and V3 conditional on the presence of their constituents. N = 1,000, default parameters.

Considering V4 to be measured, marginalizing out V5 and V6, and taking the double headed arrows to be partial deterministic, with the following parametrization:

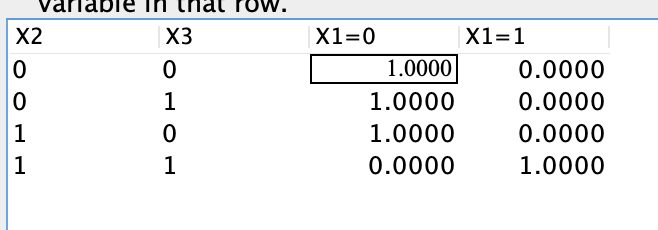
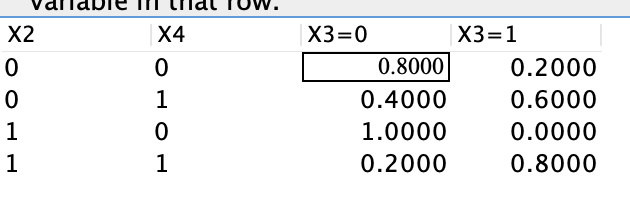
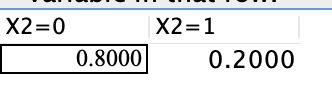
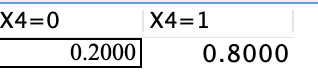


Table 1: The conditional probability table for the graph in figure 1 with X4 measured and X5 and X6 marginalized out.

The table starts with the exogenous variables, X4 and X2, which are assigned arbitrary non-extremal values. X3 is required to be almost (but see note 11) absent (0) if it’s component, X4, is absent. Otherwise, X3 is somewhat more likely to be present (1) if both its cause and its component are present. X1 is required to be absent if its components, X2 and X3 are absent, and substantially likely to be present if both its components are present. The result of the PC search using 1,000 cases randomly generated according to Table 1 is shown in Figure 3.[[16]](#footnote-16)

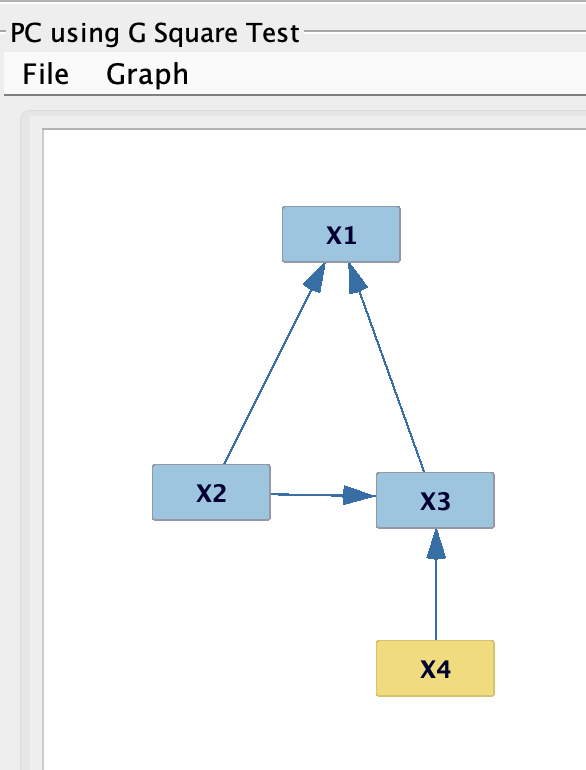


Figure 3: The PC result.

PC does not say which relations are causal and which are constitutive. If X5 and X6 are included, PC makes a hash of it.

A newer algorithm, BOSS (Andrews, et al., 2023), which uses “BN axioms” in a permutation strategy and does not require faithfulness, misses the same edge and returns Figure 3 when X4 is measured and X5 and X6 are not marginalized out.

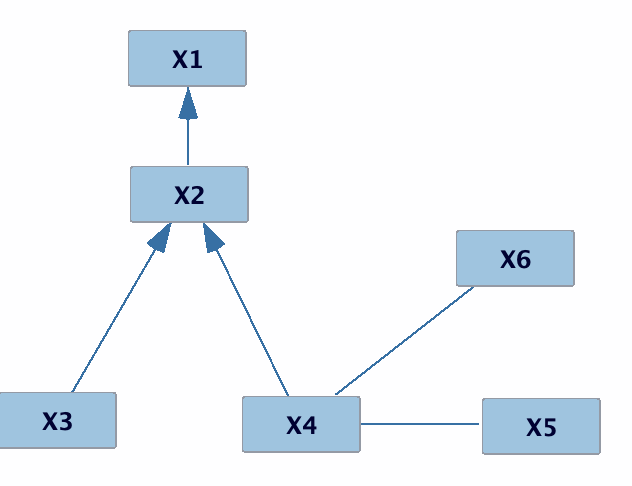
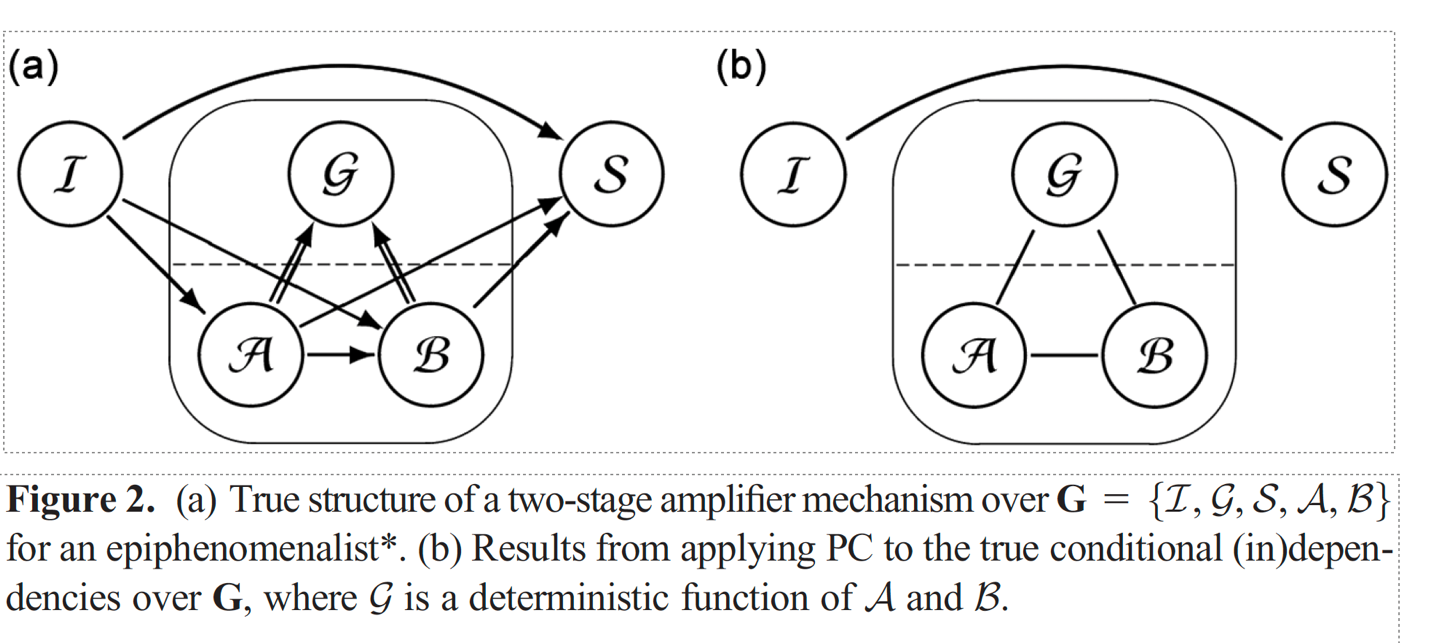


Figure 3: BOSS result on semi-deterministic data from Figure 1, N = 1,000, default parameters.

In these examples, PC and BOSS typically make an error of omission, not false positives.

B and C Example 3



G, the gain of the amplifier, is a sum of values of transistors A and B minus a base value. Conditioning on G and B makes A constant, and hence independent of I and S, and symmetrically conditioning on G and A.

B and C are correct. Quite generally, PC analysis for a set of variables can be defeated by introducing a novel variable that is a deterministic function of other variables and conditioning on the novel variable or on that variable and one or another of the variables of which it is a function. So don’t do that.

1. Corresponding author. cg09@andrew.cmu.edu [↑](#footnote-ref-1)
2. My italics. “BN axioms” refer to constraints characterizing a subclass of directed acyclic graphs whose vertices are random variables paired with a joint probability distribution on those variables subject to two well-known constraints, the Markov Condition and the Faithfulness Condition. “PC “refers to the original correct algorithm for finding sets of graphs that satisfy these (and some other) conditions from samples from a probability distribution. [↑](#footnote-ref-2)
3. More exactly, to marginalize out variables that are defined as functions of other measured variables. [↑](#footnote-ref-3)
4. Our italics. [↑](#footnote-ref-4)
5. We address the examples in B and C’s paper in Appendix 2. [↑](#footnote-ref-5)
6. In the sense that PC discovers the set of all graphs that imply the same conditional independence relations as the true graph given true information about conditional independence., and in the sense that PC is pointwise consistent in the statistical sense, i.e., with i.i.d. data it converges (pointwise) in probability to the Markov Equivalence Class of the true DAG. [↑](#footnote-ref-6)
7. We use TETRAD 7.6.2, available on GitHub. [↑](#footnote-ref-7)
8. The modification was to omit the second stage of PC that directs edges found in the first stage. [↑](#footnote-ref-8)
9. Since the frequencies measured in the JHU and JPL libraries are not exactly the same, before the experiment Roush did some interpolation to match JHU frequencies with those in JPL. Basically, a JHU frequency was mapped to the closest JPL frequency--a tiny distance to the right or left of the spectrum. [↑](#footnote-ref-9)
10. A likely explanation is that Roush was expert at identifying carbonate minerals with which he had a lot of prior experience. Almost all naturally occurring carbonate deposits on Earth are calcite or dolomite. Roush was interested in whether his identification failures were specific to him, so he hired a new spectroscopist to repeat the JHU mineral identification test. The new spectroscopist refused and quit NASA. [↑](#footnote-ref-10)
11. Rawlings used a causal inference problem as the motivating example throughout his book, although in the end he admitted failure. Later, the salient causal questions in the example were answered by the PC algorithm using Rawling’s published data summary. The PC answers were confirmed by a greenhouse experiment (Spirtes, et al, 1993). Many textbooks warn against using regression for causal inference but then use it to that end in examples and exercises. [↑](#footnote-ref-11)
12. The true graph was sparse—of average degree 2—which was essential for PC to run in a reasonable time, i.e., 12 hours on the Pittsburgh Supercomputer. [↑](#footnote-ref-12)
13. Bas van Fraassen, to CG. [↑](#footnote-ref-13)
14. Thanks to Michael Baumgartner for sharing model parameters, and to Alexander Gebharter, Conor Mayo-Wilson James Woodward, and Jiji Zhang for reading drafts. [↑](#footnote-ref-14)
15. Projection of a continuous variable onto values of a categorical variable is a different matter. So are techniques of piecewise regression that divide a continuous variable into segments and separately analyze the relation of each piece with an outcome variable, ignoring all other pieces. [↑](#footnote-ref-15)
16. It is critical to use the right statistical tests according to the variable types. Binary variables require a discrete test such as G2. If 0, 1 are converted to 0.0 or 1.0 respectively, or 7.6.2, available in GitHub. We used the default settings of search parameters. [↑](#footnote-ref-16)