

Why the Increasing Use of Complex Causal Models Is a Problem: On the Danger Sophisticated Theoretical Narratives Pose to Truth

Organizational Research Methods
1-14

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1094428119893452

journals.sagepub.com/home/ormRohny Saylor¹  and David Trafimow²

Abstract

Causal models in organizational research are complex. As use of complex models increases, the joint probability a published model is true decreases. Across *The Academy of Management Journal (AMJ)*, *Organizational Behavior and Human Decision Processes (OBHDP)*, and *Administrative Science Quarterly (ASQ)* from 2016 to 2018, it was most common to see six variables in a causal model. Even with a generous 80% independent probability of each correlation being properly theorized, the joint probability of a six-variable model is about 3.5%. Further, causal models often involve a causal chain, rendering the model even more improbable. Consequently, much of the knowledge generated in top journals is likely false. We explain that peer review demands for sophisticated theoretical narratives may pressure researchers to produce models that are embarrassingly unlikely. Traditionally, researchers argue that a low probability model is overcome by prior theory. Using an ethnostatistical Bayesian analysis, we found that given a generous prior likelihood ratio of 20, the posterior likelihood ratio is less than 1. Finally, we add “not reporting belief in a complex model” to the domain of questionable research practices and discuss auxiliary assumptions, the unstated assumptions that contextualize a theory. To ease reporting on belief in a complex model please see the following calculator: <https://practiceoftheory.weebly.com/a-causal-models-probability-of-being-true.html>.

Keywords

philosophy of science, quantitative, structural equation modeling, reliability, validity, research design, generalizability theory

To invent without scruple a new principle to every new phenomenon, instead of adapting it to the old; to overload our hypotheses with a variety of this kind; are certain proofs, that none of

¹Washington State University, Vancouver, WA, USA

²New Mexico State University, Las Cruces, NM, USA

Corresponding Author:

Rohny Saylor, New Mexico State University, 14204 NE Salmon Creek Ave VCLS 308-B, Vancouver, WA 98686, USA.

Email: r.saylor@wsu.edu

these principles is the just one, and that we only desire, by a number of falsehoods, to cover our ignorance of the truth.

—David Hume (1739, *A Treatise of Human Nature*, Part 1, Section 3, *Of Pride and Humanity*)

Presently, organizational research is rife with complex models. Editors and reviewers often see the introduction of new mediators and moderators as adding novelty or sophistication (Antonakis, 2017; Davis, 1971). However, elongating causal chains leads to diminishing likelihoods of model truth, model truth being crucial to the goals of quantitative organizational scholars. Associated problems in organizational research include questionable research design practices (Banks et al., 2016), irreproducibility (Open Science Collaboration, 2015), and disjointed claims to knowledge (Antonakis, 2017). The intent of this article is to build a new consensus around avoiding elongated causal chains. We propose raising awareness of the problem by addressing justified belief in complex models. To enable us to raise awareness of this problem, we have produced an easy-to-use web page where researchers and reviewers can calculate the probability of their causal model being true (To ease reporting on belief in a complex model please see the following calculator: <https://practiceoftheory.weebly.com/a-causal-models-probability-of-being-true.html>; see online Appendix 1 for an example).

Even a superficial perusal of top journals in organizational research renders salient the complexity and number of the causal models represented therein. For example, between 2008 and 2017, the number of articles in *The Academy of Management Journal* (*AMJ*) using causal arrows to connect variables went from 22% of articles to 36% of articles, a 64% increase; *Organizational Behavior and Human Decision Processes* (*OBHDP*) went from 24% to 60%, a 150% increase; and *Administrative Science Quarterly* (*ASQ*) went from 13% to 19%, a 46% increase. Additionally, from 2016 to 2018, the average number of variables connected by causal arrows in *AMJ*, *OBHDP*, and *ASQ* were 5.7, 5.4, and 7, respectively. Further, from 2016 to 2018, 10 articles from *AMJ*, *OBHDP*, and *ASQ* were more complex than the most complex model from any of those journals in 2008. Specifically, the 10 most complex models from 2016 to 2018 averaged 16 variables, whereas the most complex model in 2008 had 12. As the number of correlations in the matrix increases, the probability that all of them have the necessary causal status demanded by the causal model's theory decreases. This is critical because as the number of variables in the model increases, the likelihood that the model is true decreases rapidly. Of course, there are factors that mitigate the problem, such as the role that theory plays in supporting causal models or the role played by prior empirical evidence. Eventually, we will discuss both the issue of probabilities and the issue of mitigating factors.

Before speaking to these issues, however, it is first necessary to gain an appreciation of the relation between the number of variables in a model and the number of correlations in the correlation matrix that underlies whatever type of causal analysis the researcher performs to support the model. When there are only two variables in the causal model (A causes B ; or $A \rightarrow B$), there is only a single correlation coefficient underlying the causal analysis (r_{AB}). When there are three variables in the causal model (A causes B causes C ; or $A \rightarrow B \rightarrow C$), there are three correlation coefficients underlying the causal analysis (r_{AB} , r_{BC} , r_{AC}). Advancing to the case where there are six variables, there are 15 correlation coefficients underlying the causal analysis. Finally, in the past 3 years, *AMJ*, *OBHDP*, and *ASQ* have published causal models with a maximum of 16, 21, and 32 variables, representing 120, 210, and 496 correlations, respectively. For reference, for the smallest of these, a 99% probability of every link being properly theorized would lead to a 30% chance that the whole is true, and for the largest, the probability is 0.007%. Figure 1 illustrates the rapid increase in correlations and thus rapid returns on parsimony a single variable can yield. When we refer to variables, we refer to the variables theorized as causal in a model. For example, in structural equation modeling (SEM), we count latent variables rather than manifest indicators. Similarly, we do not count control variables unless they are in the causal model theorized by the researchers. We only count quantitative articles because qualitative work often theorizes processes, not causes, by using descriptions

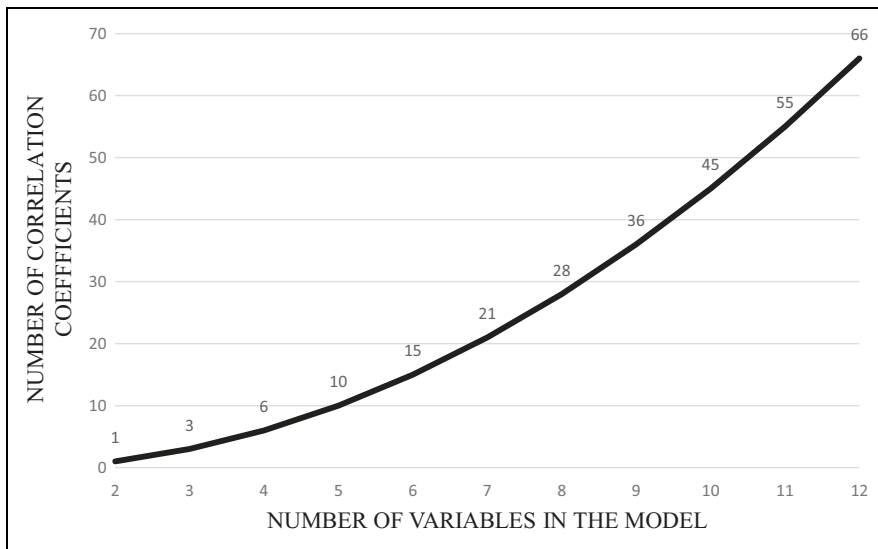


Figure 1. The number of correlation coefficients is represented along the vertical axis as a function of the number of variables in the causal model.

that share meaning by implying more than can be said (Saylor et al., 2014). Finally, our arguments concern theorized causal chains, not specific statistical methods like regression or SEM. This is because statistical methods do not change the mathematics underlying the joint probability of proper theorizing. Calculating the probability of any given model link is outside this article's scope.

Probability and Causal Status

Researchers, reviewers, and editors usually see the addition of mediators and moderators as improving the sophistication and novelty of a paper. Further, scholars tend to be less likely to believe simple narratives than complex narratives (Shepherd & Suddaby, 2017) even though the more complex a narrative becomes, the less likely it is that it will be true (Tversky & Kahneman, 1983). Thus, although a benefit of a complex theoretical narrative is increased interest (Davis, 1971), sophistication (Sutton & Staw, 1995), and excitement (Weick, 1995), the price is increased complexity with the underappreciated consequence of imperiling model truth.

Problems With Complex Causal Design

Consider the case of a simple causal model ($A \rightarrow B$) that is accompanied by a single correlation coefficient (r_{AB}). For the model to be true, the single correlation coefficient must be causal in just the right way; the correlation has to have resulted from A causing B and not from B causing A or from some outside variable causing both A and B . Put another way, the correlation between A and B has to have the right *causal status* for the $A \rightarrow B$ model to be true. Following Trafimow (2017), let us denote π_1 as the probability that r_{AB} is due to A having caused B . To move in the direction of increased complexity, let us now consider a more complex causal model, one where A causes B causes C or, using mediation language, B fully mediates the relationship between A and C ($A \rightarrow B \rightarrow C$). Now there are three correlations in the correlation matrix that underlies the causal analysis: r_{AB} , r_{BC} , and r_{AC} . Let us denote π_2 as the probability that r_{BC} is due to B causing C .

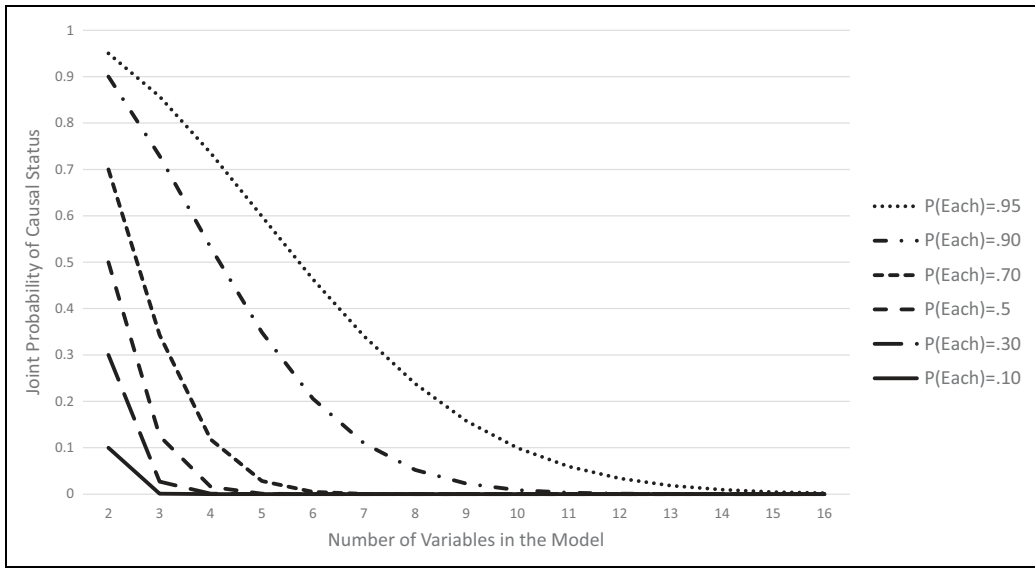


Figure 2. The joint probability that all the correlation coefficients in the underlying correlation matrix are properly theorized in terms of their causal status is represented along the vertical axis as a function of two other considerations. First, represented along the horizontal axis, is the number of variables in the causal model. Second, represented as different curves are the base probability that each correlation coefficient in the underlying correlation matrix has the appropriate causal status. Even at 95% probability for each correlation coefficient, the joint probability of the model being true falls below 50% at six variables.

We can also denote π_3 as the probability that r_{AC} is due to A causing C . In this last case, because the model is theorized so that B , rather than A , is the immediate cause of C , it would be inconvenient for the model if π_3 were a large value. It would be much more convenient if $1 - \pi_3$ were a large value. The probability that all of the correlation coefficients that pertain to the causal model have the appropriate causal status is $\pi_1\pi_2(1 - \pi_3)$. Remembering that this probability, in the context of the simple model ($A \rightarrow B$), is π_1 , it follows, ceteris paribus, that the more complex model ($A \rightarrow B \rightarrow C$) is less likely than the simple model to be true: $\pi_1 > \pi_1\pi_2(1 - \pi_3)$. We might also denote all other noncauses of C , like the unmentioned D , E , and F . And, indeed, this would be essential to include if we sought to develop theory as complex as organizational reality. However, we are advocating theory much less complex than reality and locating some complexity in the auxiliary assumptions of the theory. Consequently, and for all our analysis, we limit ourselves to the variables of theoretical interest in a causal model.

For example, it would be convenient for the mediation model if π_1 and π_2 were large values, so let us set these at .75. And it would be convenient if π_3 were a small value, so let us set this at .25. From this, we can calculate the probability that all of the correlation coefficients have the appropriate causal status: $.75 \times .75 \times (1 - .25) = 42.2\%$. Thus, even starting with favorable probabilities regarding the causal status of each of the correlation coefficients in the underlying correlation matrix, the probability of a simple three-variable causal chain is less than the probability of winning a coin flip.

As Figure 1 illustrates, as the number of variables in the model increases, the number of correlation coefficients in the underlying correlation matrix increases even more, and for the causal model to be true, all these correlations have to have the appropriate causal status. How likely is it that this is

true *ceteris paribus*? Figure 2 illustrates the problem with having complex causal models with many variables. Figure 2 represents the probability that all relevant correlation coefficients have the appropriate causal status to support the causal model along the vertical axis as a function of two other factors. The first, along the horizontal axis, is the number of variables in the causal model. We let this vary from 2 to 16, the latter value being the average for the top 10 most complex articles in *AMJ*, *OBHDP*, and *ASQ* from 2016 to 2018. The second, represented by a set of curves, is the probability that each particular correlation coefficient has the appropriate causal status. We let this base probability vary from .1 to .95.

Before continuing, it is important to avoid a potential confusion. Specifically, the probability that an obtained correlation coefficient is theorized for the “right” reason is not the same thing as the correlation itself. For example, imagine that a researcher obtains a strong correlation of .80 between *A* and *B*, thereby indicating a strong predictive relationship between *A* and *B*. This does not mean the correlation is theorized for the right reason. For example, the calculated .80 does not indicate that the arrow goes in the right direction, that there is no exogenous third cause, or that all other arrows or absence of arrows to or from *A* or *B* are properly theorized. Even more striking is the fact that actual models often involve a causal chain, making it even more improbable that the causality in the model is properly theorized.¹ Thus, the numbers we present represent the upper bound of the probability of the model being true.

It is instructive to consider the bottom quartile case that we have seen across *AMJ*, *OBHDP*, and *ASQ*, where there are just four variables and, hence, six correlation coefficients of relevance. Figure 2 shows that when the base probability that a correlation coefficient has the appropriate causal status is .5, the joint probability that all six of them have the appropriate causal status is only .016. When there are six variables, the joint probability is nearly zero. Even when we start with the extremely optimistic base probability of .95, the resulting joint probability at six variables is .46, less than a coin flip. If there is a well-validated theoretical model that has long stood the test of time, a theory can be developed that is internally consistent and no longer meets the assumption of independent probability presented here. But it is rare in the organizational sciences that we develop simple, elegant, and regularly replicated tests of theories. Instead, we seem obsessed with neophilia (Antonakis, 2017; Rhodes & Pullen, 2010), where we would prefer ever elongated causal chains to production and establishment of true theories.

Some may argue that causal models are not seeking a true model but the best model so far. However, if a model with only six variables goes from a 10% chance of each causal relationship being properly theorized to a 50% chance, the joint probability of the model being true goes from 1×10^{-15} to 3×10^{-5} . While the second model is more than 10 orders of magnitude better than the first, the actual probability of having a true model remains effectively zero. Thus, the one thing that can be said about such a model is that it is almost certainly not true. In short, demands for sophisticated theoretical narratives pose a major danger to truth!

Problems With Dependence

Thus far, we have assumed that the probability that each relevant correlation coefficient has the appropriate causal status is independent of the probability that each other relevant correlation coefficient has the appropriate causal status. Which is to say, we are not assuming that the variables are independent or even that the correlation coefficients are independent. Rather, we are assuming that the probability of the appropriate causal status of each correlation coefficient is independent of the probability of the appropriate causal status of each other correlation coefficient. We do not have to assume such independence, but we believe that independence actually provides the most optimistic scenario for the causal modeler. To see why, consider the example of four variables in the causal model so that there are six relevant correlation coefficients. Suppose that the probability that

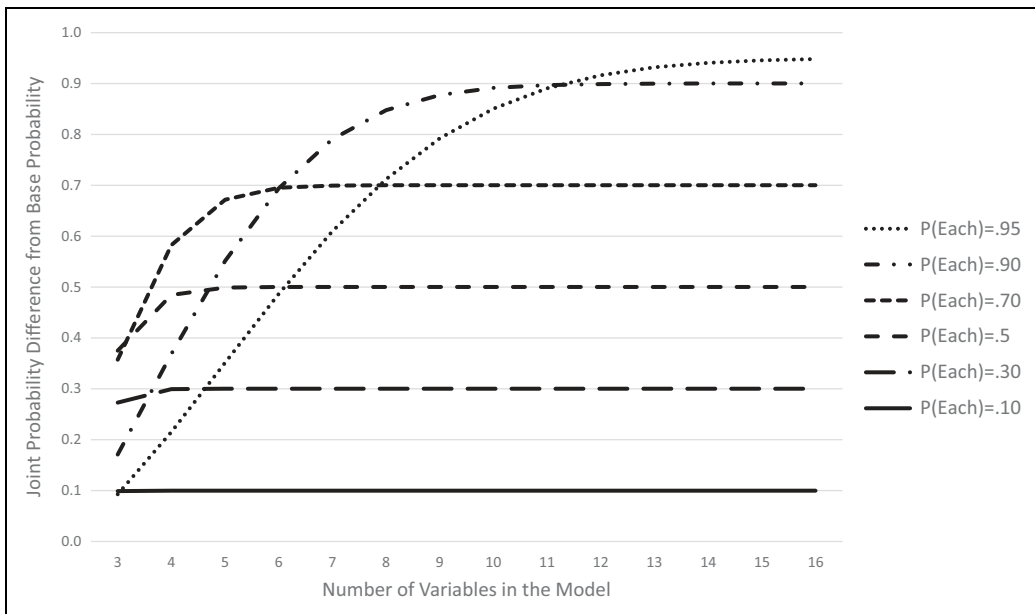


Figure 3. The difference in joint probabilities is represented along the vertical axis as a function of the number of variables along the horizontal axis.

the sixth correlation coefficient has the correct causal status depends on the other five probabilities of correct causal status. For example, we might suppose, using an optimistic case in Figure 2, that the base probability is .9. This would mean that under independence, the probability of the sixth correlation is also .9. But if we assume dependence, the probability of the sixth correlation could go up to 1, a maximum possible gain of .1; but it also can go down to 0, a maximum possible loss of .9. This problem could be mitigated by assuming a less favorable base probability so that there is more room for dependence to be helpful and less room for dependence to be harmful; but in that case, as Figure 2 shows, the joint probability suffers dramatically. Thus, the basic message of Figure 2 is exacerbated rather than mitigated by assuming dependence.

Simple Truth

We now investigate the difference between (a) a model with maximal simplicity and (b) the probability that all of the relevant correlation coefficients have the appropriate causal status. This difference in joint probabilities is illustrated in Figure 3 along the vertical axis. Most generally, Figure 3 shows that the simplest possible causal model, involving two variables and one relevant correlation coefficient, is much more likely to be true than the types of complex models that litter the organizational research landscape. Like Figure 2, Figure 3 includes the number of model variables along the horizontal axis and curves representing base probabilities ranging from .1 to .95. Figure 3 shows that the difference in joint probabilities becomes extreme with only a slight increase in complexity and reaches asymptote quickly, with the most exaggerated effect occurring when the base probability is .95 and the smallest effect occurring when the base probability is .1. A comparison between Figures 2 and 3 places the causal modeler in an interesting dilemma. That is, to minimize the foregoing effect, it is necessary to have a low base probability. But if one starts with a low base probability, Figure 2 shows that the resulting joint probability will be unacceptably low.

Statistical Indistinguishability and Mitigating Factors

An important problem with any sort of causal analysis based on a correlation coefficient or set of correlation coefficients is that many possible models are consistent with the analysis (e.g., Fiedler et al., 2011; Grice et al., 2015; Kline, 2015; MacKinnon et al., 2000; Tate, 2015; Thoemmes, 2015; Trafimow, 2015). As Spirtes et al. (2000) stated, “Without experimental manipulations, the resolving power of any possible method for inferring causal structure from statistical relationships is limited by statistical indistinguishability. If two causal structures can equally account for the same statistics, then no statistics can distinguish them” (p. 59).

Sophisticated causal modelers are aware of the issue of statistical indistinguishability and have attempted to counter it in two ways. One way is to argue that path models are not really “causal” but merely represent relationships between variables. We believe that this argument is disingenuous. For one thing, the connectors in the models are arrows rather than line segments, thereby indicating causation rather than simply statistical relationships. In addition, the causal model is often accompanied by a practical recommendation of the nature that the findings influence how practitioners ought to engage in some course of action to cause a desired result. If the arrows in models are only representing statistical relations and are not representing causal relations, it does not make sense to recommend a course of action that is expected to cause a particular outcome. We might add that writers of regression books have commented on the disingenuous claim that arrows in path diagrams are not supposed to represent causation and have avowed that the arrows are in fact supposed to represent causation (e.g., Cohen et al., 2003; McClendon, 1994; Pedhazur, 1997).

With disingenuousness out of the way, let us consider the more common and intellectually more honest claim that theory makes up for the statistical indistinguishability problem. That is, although the causal analysis itself may suffer from the statistical indistinguishability problem, the problem is mitigated by the presence of strong theory. We certainly agree that data are better when derived from strong theory rather than absent theory. However, we question whether causal models in top organizational research journals really are derived from strong theory. Weick (1995) pointedly argued that causal models in the organizational sciences are not derived from basic theories. Rather, such causal models are derived from a combination of previous findings, scattered intuitions, and of course, the obtained data. In this case, it seems circular to say that theory supports the causal model when it is really the causal model that supports the theory.

Sophisticated causal modelers may suggest we distinguish between the causal model without path coefficients versus the causal model with path coefficients. The “abstract” path model that does not yet have empirically obtained path coefficients accompanying the arrows might be characterized as being at the “theoretical level,” whereas the same model accompanied by empirically obtained path coefficients might be characterized as being at the “empirical level.” To the extent that the theoretical and empirical models match, the empirical causal model supports the theoretical causal model. However, there are practical problems with this philosophical take. The first problem with comparing theoretical to empirical models is that the support that the empirical model makes to the theoretical model is weakened by the statistical indistinguishability problem as stated previously. In addition, the empirical model depends on the causal statuses of the correlation coefficients on which the causal analysis is based. And as we have seen, as complexity increases even a little, the joint probability decreases rapidly.

How important are these problems concerning theory and data in the context of the issue of complexity? It is possible to answer this question by using a Bayesian perspective. We might consider the joint probabilities graphed in Figure 2 to be Bayesian *priors*, to be updated due to mitigating factors such as theory and causal analysis. With this in mind, let us consider the famous theorem by Bayes.

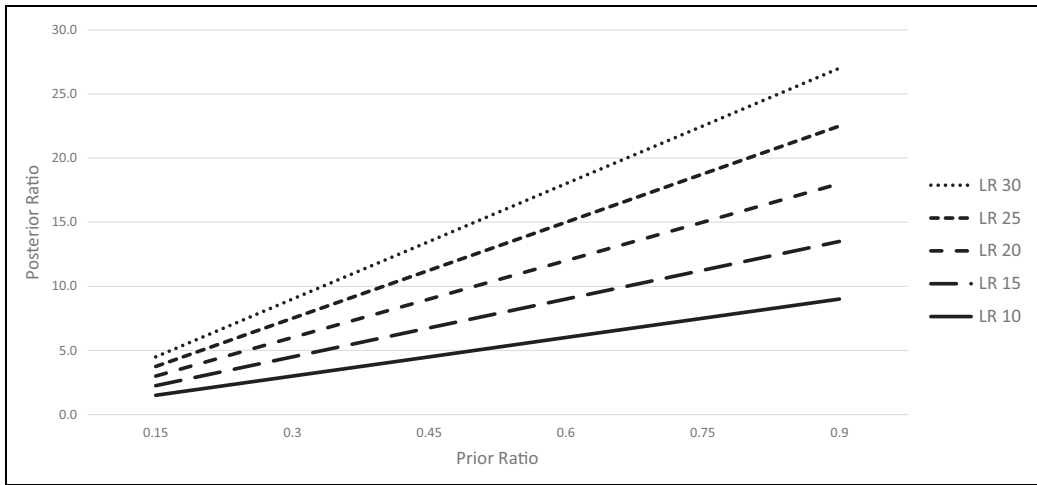


Figure 4. The posterior ratio is represented along the vertical axis as a function of the prior probability of the model along the horizontal axis and the likelihood ratio. While the calculation of likelihood ratios is outside of the scope of this article, note that even the field of astronomy, with its direct observation of physical phenomenon, often uses a likelihood ratio of 20 [i.e. $1/(1-.95)$] as the highest reasonable likelihood ratio.

The Bayesian Way

A key ethnostatistical method (Gephardt & Saylor, 2019) for addressing questions of quantitative knowledge production is the use of simulation. Here we do this in a Bayesian way. Consider a causal model with however many variables and number of underlying correlations. From a Bayesian perspective, we are interested in updating the prior probability of the model ($P(M)$) based on the probability of the data given the model ($P(D|M)$) and the probability of the data ($P(D)$). Equation 1 gives the classic Bayesian answer for the posterior probability of the model, which also is the probability of the model given the data ($P(M|D)$).

$$P(M|D) = \frac{P(M)P(D|M)}{P(D)} \quad (1)$$

Equation 2 gives the Bayesian answer for the probability that the model is not true (\sim) given the data.

$$P(\sim M|D) = \frac{P(\sim M)P(D|\sim M)}{P(D)} \quad (2)$$

Dividing Equation 1 by Equation 2 gives the ratio form of Bayes's theorem as it applies to the probabilities of the model being true or not true. The three ratios in Equation 3 can be termed the posterior ratio (P_oR), the prior ratio (P_rR), and the likelihood ratio (LR), respectively.

$$\frac{P(M|D)}{P(\sim M|D)} = \frac{P(M)}{P(\sim M)} \frac{P(D|M)}{P(D|\sim M)} = P_oR = P_rR \times LR \quad (3)$$

To illustrate the implications of Equation 3, Figure 4 expresses the posterior ratio along the vertical axis as a function of the prior ratio along the horizontal axis (varying from .3 to .9) and with likelihood ratios of 10, 15, 20, 25, and 30. Note that, for example, if the LR is 10, it means that the data are 10-fold more likely given that the model is true than if the model is not true. Thus, even the lowest LR we included in Figure 4 can be argued to overstate the quality of most data (e.g.,

Fiedler et al., 2011; Grice et al., 2015; Kline, 2015; MacKinnon et al., 2000; Tate, 2015; Thoemmes, 2015; Trafimow, 2015). And the largest likelihood ratio of 30 wildly overstates of the quality of any such data. Nevertheless, Figure 4 results in pessimistic conclusions regarding the posterior probability of organizational research models, as we discuss in the following.

Recall again that the joint probability in Figure 2 for a model with six variables and assuming an unrealistically optimistic base probability of .9 was .21, which makes the prior ratio² .27. If we approximate the prior ratio at a favorable .3 for comparison in Figure 4, we see that at an LR of 30, the posterior ratio is 9; for an LR of 20, the posterior ratio is 6; and for an LR of 10, the posterior ratio is 3. Is this a convincing posterior ratio? Bayesians have suggested different levels for what counts as a convincing posterior ratio, but there has been a reasonable amount of convergence on 10 as a minimum bar (Etz & Vandekerckhove, 2016). Thus, in the present case, and where the data are 30 times more likely given that the causal model is true than given that the causal model is not true, the posterior case for the modal model still does not pass the bar. Let us now consider the influence of reducing the size of the model by a single variable. Using the same optimistic base probability of .9 but five variables rather than six leads to joint probability of .35, thus making the prior ratio .54. If we again approximate this prior ratio favorably at .6 for comparison in Figure 4, we see that an LR of 30 has the posterior ratio of 18, an LR of 20 has a posterior ratio of 12, and an LR of 10 has the posterior ratio of 6. In our judgment, Figure 4 presents an extremely optimistic range of values. For example, for the more realistic yet still generous base probability of .7 and the simple four-variable model, crossing the posterior ratio bar of 10 requires an LR in excess of 75. At the more realistic yet still generous LR of 10, the posterior ratio is less than 1.4. Obviously, this value fails to reach the cutoff, and the posterior ratios are even lower when we assume less optimistic likelihood ratios or start with less optimistic base probabilities. Because prior belief in a model has a large influence over whether the model is questionable, we add “not reporting belief in a complex model”³ to the domain of questionable research practices (Banks et al., 2016).

The Potential of Auxiliary Assumptions + Simple Theory

Despite the present demonstrations, there are arguments that seemingly favor complex causal models. Reality is complex, thereby leading many researchers to believe that complex models are needed to reflect that reality. There are also issues pertaining to playing the academic game. Organizational research journals are more likely to publish complex than simple models. One reason for this is the focus that journal editors have on novelty. If a model is sufficiently complex, it is unlikely to have been published before, at least not in its entirety, thereby justifying that there is a novel contribution. Yet another reason is the issue of alternative explanations. To see this, imagine that a researcher submits a manuscript featuring a simple zero-order correlation coefficient. The reviewers would immediately point out that one variable could have caused the other, the other could have caused the first, or some outside variable could have caused them both. Smart reviewers would have no trouble coming up with plausible outside variables. In contrast, for a complex model, it takes too much cognitive work for reviewers to find alternative explanations for all the pathways, and so sophisticated theoretical narratives are much more likely to wend their way through the review process.

Because the foregoing demonstrations show that complex models are much less likely to be true than simple models, let us return to the issues of the complexity of reality and novelty. There can be little doubt about the complexity of reality; however, that said, the complexity of reality does not necessarily entail that theories need to be equally complex. The obvious example of this is Newtonian theory in physics, which depends on a few basic principles. These are the three laws of motion and the law of gravity. As Cartwright and McMullin (1984) documented, as a literal description of reality, Newtonian theory is simply wrong. The simplicity of the theory is unable to perfectly account for reality. However, the ability of a theory to account for reality is not the best way to

judge its worth because the ability of a theory to account for reality depends on, to an underappreciated extent, auxiliary assumptions.

The Importance of Auxiliary Assumptions

To defend the foregoing claim of how important auxiliary assumptions are, it may be easiest to start with the falsification issue popularized by Sir Karl Popper (Lakatos, 1976). Popper's reasoning rests on the invalidity of the following syllogism involving a theory and a confirmed prediction:

- If the theory is true, then the prediction should work;
- the prediction works;
- therefore, the theory is true.

The syllogism is blatantly invalid because the prediction could have worked for some reason other than the theory. This is the logical fallacy called affirming the consequent. Thus, at best, confirming the prediction in a study can only support the theory but cannot prove it. In contrast, there is another syllogism that is logically valid and provided the basis for Popper's thinking:

- If the theory is true, then the prediction should work;
- the prediction does not work;
- therefore, the theory is not true.

As the syllogism makes clear, disconfirming a prediction logically disproves the theory, at least from the point of view of strict logic. Therefore, Popper advocated that researchers attempt to disprove rather than prove theories and that science advances by replacing disproved theories with better ones. But alas, strict logic is not the only factor. As many have pointed out (e.g., Lakatos, 1976; Trafimow, 2009), scientists do not arrive at predictions merely from theories, and for a very good reason. Theories depend strongly on nonobservational terms, whereas empirical predications consist of observational terms. To stay with Newton, consider his famous equation: *force = mass × acceleration*. As Nobel Laureate Leon Lederman (1993; Lederman and Hill, 2004) pointed out, mass is a nonobservational term that Newton never defined independently of the other constructs in the equation. Nor have physicists ever succeeded in independently defining it. Yet, mass is a key concept in Newton's theory and in physics more generally. If a scientist were to make a prediction with respect to mass, such as how much Churchill's, *A History of the English-Speaking Peoples* would weigh on the moon, the scientist would need to make auxiliary assumptions not in Newton's theory: about characteristics of the moon, characteristics of the measuring device, which version of Churchill's work is to be used, and so on. Stated more philosophically, to bring the nonobservational term, *mass*, down to the level of an observational term, *weight*, it is necessary to make assumptions not contained in Newton's theory. Auxiliary assumptions are often tacit and thus in need of testing (Gray & Cooper, 2010).

This situation is not unique to physics, of course. Consider the theory of planned behavior (Ajzen, 1991) and the many guises it takes throughout organizational research (Krueger et al., 2000; Venkatesh & Davis, 2000). Just as with mass, the elements theorized to exist within a person's mind (nonobservational terms) must be translated into observational terms. Thus, there are auxiliary assumptions linking the nonobservational terms in the theory to observational terms in empirical hypotheses.

Because there is no way to test a theory in the absence of auxiliary assumptions, the major premise in Popper's valid syllogism is incorrect, thereby rendering the syllogism logically valid but unsound because of the incorrect major premise. Let us construct another syllogism that is both logically valid and contains true premises:

- If the theory is true and countless auxiliary assumptions are true, then the prediction should work;
- the prediction does not work;
- therefore, the theory is not true or at least one auxiliary assumption is not true.

Thus, a failed prediction can be blamed on the theory; but it also can be blamed on one or more auxiliary assumptions. Because it is not clear where the blame lies, a failed prediction cannot disprove a theory any more than a confirmed prediction can prove a theory. Thus, the importance of auxiliary assumptions in the progress of science is difficult to overstate.

The Potential Power of Simple Theory

Models as complex as reality are an exercise in futility because reality cannot be fully represented by a model less complex than itself (Carroll, 1894; Korzybski, 1958). To see this, it is worth considering again that Newton's theory does not mimic reality. But if Newton's theory does not mimic reality, why do physicists agree that the theory constituted one of the most important advances in the history of physics? The answer is that Newton's theory, just like other top theories in physics, provides a description of an idealized universe. In turn, with high-quality auxiliary assumptions, researchers can make predictions about the real universe using a theory and well-chosen auxiliary assumptions. Edmund Halley provided an example by predicting the time of return of the comet that now bears his name. Halley used Newton's theory but also made auxiliary assumptions not contained in the theory regarding diverse matters such as the present position of the comet, the presence of gravitationally relevant astronomical bodies, and others.

To illustrate the distinction between auxiliary assumptions and basic theory, consider dropping a heavy book or a feather from the same height at the same time. According to Newton's theory, both objects should hit the ground at the same time. Of course, there is no way this would work because of air resistance. The heavy book would be less subject than the feather would be to the effects of air resistance; the book would hit the ground much more quickly than would the feather. One might be tempted to use the failed prediction to question the worth of Newton's theory. A proper response might be to argue that anyone who tests Newton's theory in this way is guilty of performing an inappropriate experiment. There is no air—and hence no air resistance—in Newton's theory. A researcher might design an interesting study where the experiment was performed in a vacuum. While this experiment would be further from reality, it would be closer to Newton's idealized universe and thereby address a critical auxiliary assumption. In a vacuum, the two objects really would hit the ground at close to the same time. Thus, by adroit use of auxiliary assumptions, a researcher can adapt a simple theory—such as Newton's—to make excellent predictions in the real universe. In this way, scholars have generated the study of aerodynamics and the field of aerospace engineering.

Conclusion

Our analysis shows that increasingly complex causal models are populating top organizational research journals. Further, we have demonstrated that no amount of prior theory can justify elongated chains that we see in modern organizational research. Consequently, sophisticated theoretical narratives demanded by many top journals pose a danger to model truth. Given how unlikely most causal models are, the danger is hard to understate. What an embarrassing waste of human potential it would be if organizational research devolved to nothing but the production of lists of mediators and moderators for different dependent variables! To fight this danger, we propose researchers consider the joint probability of their model being true. As a counterexample, it is interesting to

consider physics, which does not attempt to represent a complex universe using complex theory but rather uses simple theories that describe idealized universes. In contrast to organizational researchers, for physics researchers, the complexity is located in the auxiliary assumptions rather than in the theory. We recognize that any researcher who moved in this direction would immediately be accused of external invalidity. However, because we have shown that external invalidity can be a good thing if it enables the use of simple theory along with contextualizing auxiliary assumptions, such accusations need not discommode the philosophically sophisticated organizational researcher. Of course, there is also the practical matter of how such studies can wend their way through the review process, how to design such studies, and what criteria to use for such reviews. But that is a paper for another day.

Acknowledgments

We would like to thank Associate Editor Professor Louis Tay, whose guidance and support has made this article possible. Further, we would like to thank the three anonymous peer reviewers for their insights, to acknowledge Garret Cayan and Jacob Klopp for helping us with our count of variables, and to thank Jacob Klopp, Charlene Debus, and Frank Flores for offering comments on the manuscript.

Authors' Contribution

This article relies on Rohny Saylor's critical performative perspective, specifically that of ethnostatistics (Gephart & Saylor, 2019) and David Trafimow's development of critical quantitative insights (Trafimow, 2019). The initial intuition, quantitative argumentation, and math proofs were handled primarily by David Trafimow. The theoretical framing, illustrations, and leading of the paper through peer review was handled primarily by Rohny Saylor.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Rohny Saylor  <https://orcid.org/0000-0001-6263-0997>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. We must thank the associate editor, Professor Louis Tay, for this clear insight.
2. The prior ratio is a function of the prior probability of the model: $P_r R = \frac{P(M)}{1-P(M)}$. Thus, to obtain Figure 4 via Equation 3, it is merely necessary to convert from the joint probability of the model in Figure 2 to the prior ratio, and then the prior ratio can be multiplied by the likelihood ratio, as Equation 3 illustrates.
3. This link can be used by researchers to calculate the joint probability of their models: <https://practiceoftheory.weebly.com/a-causal-models-probability-of-being-true.html>.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Antonakis, J. (2017). Editorial: The future of *The Leadership Quarterly*. *The Leadership Quarterly*, 28, 1-4.

- Banks, G. C., O'Boyle, E. H., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A. A., & Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, *42*(1), 5-20.
- Carroll, L. (1894). *Sylvie and Bruno concluded*. Macmillan.
- Cartwright, N., & McMullin, E. (1984). *How the laws of physics lie*. Oxford University Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates, Publishers.
- Davis, M. S. (1971). That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, *1*(2), 309-344.
- Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the Reproducibility Project: Psychology. *PLOS ONE*, *11*(2), Article e0149794. <https://doi.org/10.1371/journal.pone.0149794>
- Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231-1236. <https://doi.org/10.1016/j.jesp.2011.05.007>
- Gephart, R. P., & Saylor, R. (2019). Etnostatistics. In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), Sage Research Foundations. <https://dx.doi.org/10.4135/9781526421036747883>
- Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods*, *13*(4), 620-643.
- Grice, J. W., Cohn, A., Ramsey, R. R., & Chaney, J. M. (2015). On muddled reasoning and mediation modeling. *Basic and Applied Social Psychology*, *37*(4), 214-225. <https://doi.org/10.1080/01973533.2015.1049350>
- Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, *37*(4), 202-213. <https://doi.org/10.1080/01973533.2015.1049349>
- Korzybski, A. (1958). *Science and sanity: An introduction to non-Aristotelian systems and general semantics*. Institute of General Semantics.
- Krueger, N. F., Jr., Reilly, M. D., & Carsrud, A. L. (2000). Competing models of entrepreneurial intentions. *Journal of Business Venturing*, *15*(5-6), 411-432.
- Lakatos, I. (1976). Falsification and the methodology of scientific research programmes. In S. G. Harding (Ed.), *Can theories be refuted?* (pp. 205-259). Springer.
- Lederman, L. (1993). *The God particle: If the universe is the answer, what is the question?* Houghton Mifflin.
- Lederman, L. M., & Hill, C. T. (2004). *Symmetry and the beautiful universe*. Prometheus Books.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173-181. <https://doi.org/10.1023/A:1026595011371>
- McClendon, M. J. (1994). *Multiple regression and causal analysis*. F. E. Peacock Publishers, Inc.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. <https://doi.org/10.1126/science.aac4716>
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Wadsworth.
- Rhodes, C., & Pullen, A. (2010). Editorial: Neophilia and organization. *Culture and Organization*, *16*, 1-6.
- Saylor, R., Boje, D. M., & Mueller, T. J. (2014). Entrepreneurial storytelling in moments of friendship: Antenarratives of business plans, risk taking, and venture capital narratives. *Tamara: Journal for Critical Organization Inquiry*, *12*(4). <https://tamarajournal.com/index.php/tamara/article/view/395>
- Shepherd, D. A., & Suddaby, R. (2017). Theory building: A review and integration. *Journal of Management*, *43*(1), 59-86.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*(3), 371-384.
- Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, *37*(4), 235-246. <https://doi.org/10.1080/01973533.2015.1062380>
- Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology*, *37*(4), 226-234. <https://doi.org/10.1080/01973533.2015.1049351>
- Trafimow, D. (2009). The theory of reasoned action: A case study of falsification in psychology. *Theory & Psychology*, *19*(4), 501-518.

- Trafimow, D. (2015). Introduction to special issue: What if planetary scientists used mediation analysis to infer causation? *Basic and Applied Social Psychology*, 37(4), 197-201. <https://doi.org/10.1080/01973533.2015.1064290>
- Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics. *Educational and Psychological Measurement*, 77(5), 831-854.
- Trafimow, D. (2019). A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics*, 7, 26.
- Tversky, A., & Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315. <https://doi.org/10.1037/0033-295X.90.4.293>
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, 40(3), 385-390.

Author Biographies

Rohny Saylor is an assistant professor of entrepreneurship at Washington State University. His research focuses on entrepreneurial storytelling processes and methods. His passion is the advancement of human creativity, hope, and authentic compassion through and within organizational scholarship. Dr. Saylor has published in *Organizational Research Methods* and *Human Relations*.

David Trafimow's interest is in social cognition and methodological, statistical, and philosophical issues pertaining to science. He has written more than 100 peer-review journal articles, many in top journals such as the *Organizational Behavior and Human Decision Processes*, *Journal of Personality and Social Psychology*, and *Psychological Review*.