Are complex causal models less likely to be true than simple ones? A critical comment on Trafimow (2017)

Are Complex Causal Models Less Likely to Be True Than Simple Ones? A Critical Comment on Trafimow (2017)

Marco Del Giudice

University of New Mexico

In press: ***Behavior Research Methods***

Marco Del Giudice, Department of Psychology, University of New Mexico.
Address correspondence to Marco Del Giudice, Department of Psychology, University of New Mexico. Logan Hall, 2001 Redondo Dr. NE, Albuquerque, NM 87131, USA; email: marcodg@unm.edu

Abstract

Trafimow (2017) used probabilistic reasoning to argue that more complex causal models are less likely to be true than simpler ones, and that researchers should be skeptical of causal models involving more than a handful of variables (or even a single correlation coefficient) [Trafimow, D. (2017). The probability of simple versus complex causal models in causal analyses. *Behavior Research Methods, 49,* 739-746]. In this comment, I point out that Trafimow's argument is misleading, and reduces to the observation that *more informative* models (that make definite statements about certain causal relations) are less likely to be true than *less informative* models (that remain silent about those relations, by omitting some variables from consideration). This correct but trivial statement does not deliver the epistemological leverage promised in the paper. When complexity is evaluated with reasonable criteria (such as the number of nonzero effects in alternative models involving the same variables), more complex models can be more, less, or equally likely to be true compared with simpler ones. I also discuss Trafimow's claim that, if a model is unlikely to be true *a priori*, researchers will seldom be able to gather evidence of sufficient quality to support it; in practice, even low-probability models can receive strong support without the need for extraordinary evidence. Researchers should evaluate the plausibility of causal models on a case-by-case basis, and be skeptical of overblown claims about the dangers of complex theories.

Keywords: Causal models; epistemology; likelihood ratio; Occam's razor; probability.

In "The probability of simple versus complex causal models in causal analyses", Trafimow (2017) used probabilistic reasoning to argue that "a simple causal model based on a single correlation coefficient is more likely to be true than a complex causal model based on several correlation coefficients" (p. 743). The author wondered: "given that a simple causal model is much more likely to be true than is a complex one, why do journal editors and reviewers favor complexity?" (p. 743). After suggesting that the answer may lie in the cognitive limitations and biases of scientists (e.g., the conjunction fallacy), he concluded that, to be rational, "researchers should be more open to simple causal models based on a single correlation coefficient *or* they should be less open to complex causal models based on many correlation coefficients" (p. 745; emphasis in the original).

Even if a given model is unlikely to be true *a priori*, the evidence in its favor can be so strong that it leads to a high posterior probability. Trafimow (2017) considered this Bayesian argument, but noted that the likelihood ratios required to overturn unfavorable prior probabilities can be large. For example, a model with a 25% prior probability of being true would require a likelihood ratio (LR) of 3 to reach a posterior ratio of 1 (the "point of indifference"), and a LR of 30 to yield a posterior ratio of 10 (an often-used conventional threshold for "strong support"). The author argued that researchers will generally be unable to collect sufficient evidence to meet this threshold: "although this may be possible, collecting data of such quality constitutes a difficult challenge for researchers to overcome" (p. 745). More recently, Saylors and Trafimow (2020) built on this argument to make the startling claim that, since most published causal analyses involve more than a handful of variables, "much of the knowledge generated in top journals is likely false" (p. 1). These authors also provided an online calculator that computes the maximum *a priori* probability that a given causal model is true, based on the number of variables involved (https://practiceoftheory.weebly.com/a-causal-models-probability-of-being-true.html).

On the surface, this is a seductive argument. It appears to formalize Occam's razor in the domain of causal models, and promises to yield tremendous leverage based on nothing more than basic laws of probability. But a closer look shows that the argument is much weaker than it appears—in fact, so weak as to be effectively useless.

The crucial flaw lies in Trafimow's notion of model complexity, which he defines as the number of variables (and corresponding number of correlations) involved in a causal model. As a result, the paper does not prove the interesting point that *simpler* models are more likely to be true, but only the trivial point that *less informative* models are more likely to be true, owing to the fact that they say nothing at all about certain variables and causal relations. Consider the toy example presented in the paper, in which the model A $\rightarrow$ B is compared with the model A $\rightarrow$ B $\rightarrow$ C and found more likely to be true (except in special cases). Importantly, the former model does *not* imply that A and C are causally unrelated, but instead leaves the question entirely open. If for example A = drinking coffee, B = focused attention, and C = learning, the A $\rightarrow$ B $\rightarrow$ C model implies that drinking coffee affects attention, which in turn affects learning. In contrast, the A $\rightarrow$ B model implies that drinking coffee affects attention, which in turn *may or may not* affect learning. By any reasonable standard, this is not just a "simpler" account of the same phenomenon; instead, the A $\rightarrow$ B model remains agnostic about the effects of coffee and attention on learning. Because it makes fewer definite claims about reality, this model is less informative than the A $\rightarrow$ B $\rightarrow$ C model.

The fact that less informative models are *a priori* more probable is obviously true—but not very useful as an epistemological tool. Consider the A → B → C hypothesis that drug A kills pathogen B, and as a result cures disease C. Naturally, this hypothesis is less likely to be true than the alternative A → B hypothesis that drug A kills pathogen B, which *may or may not* cure disease C; not because the latter is "simpler", but because it remains silent about some (important) causal relations. The same goes for the B → C hypothesis that killing pathogen B cures disease C, while drug A *may or may not* kill the pathogen. Researchers trying to find a cure for the disease may be excused if they prefer to test the more informative model, despite its lower probability.

When researchers evaluate alternative accounts of a phenomenon, they are typically interested in models that involve the same variables, and hence are equally informative about the presence vs. absence of relations among those variables. Model complexity can then be evaluated based on features such as the number of free parameters to estimate from the data. In typical applications, each additional nonzero effect (i.e., an arrow in the causal diagram) corresponds to an additional free parameter, which makes the number of effects a reasonable index of complexity. Consider the three causal models shown in Figure 1. In Model 1, A has a direct effect on C, while B is unrelated to both A and C. In Model 2, the effect of A on C is fully mediated by B. In model 3, A affects C both directly *and* via the mediating effect of B. The three models involve the same variables, but Model 2 is more complex than Model 1, and Model 3 is more complex than Model 2. Also note that both Model 1 and model 2 are nested within Model 3. This is relevant because Trafimow (2017) specified that his argument applies unambiguously only to cases in which the more complex model "subsumes" the simpler one (p. 744).

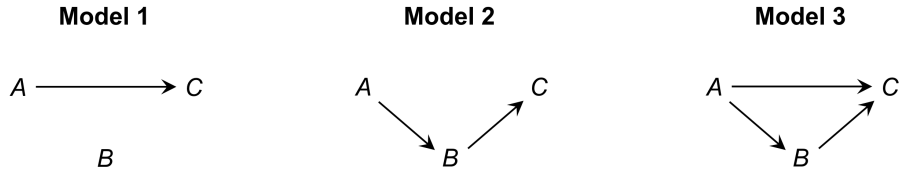

Figure 1. Three causal models of increasing complexity involving the same variables.

Following the notation in Trafimow (2017), π_1 is the probability that A → B, i.e., there is a nonzero causal relation between A and B *and* the effect runs in the hypothesized direction (from A to B); whereas $(1 - \pi_1)$ is the probability that there is no direct causal relation between A and B (Trafimow, 2017, p. 740). Thus, the example implicitly assumes that the direction of the effect (if it exists) is always correctly specified. For the sake of simplicity, in the following I make the same assumption as it does not change the substance of my critique. The other probabilities are π_2 (the probability that B → C) and π_3 (the probability that A → C). Assuming independence, the probability of Model 1 is $(1 - \pi_1)(1 - \pi_2)\pi_3$; the probability of Model 2 is $\pi_1\pi_2(1 - \pi_3)$; and the probability of Model 3 is $\pi_1\pi_2\pi_3$. It is easy to verify that, depending on the values of π_1, π_2, and π_3, *any* of the three models can become the most probable in the set. For example, Model 1 is the most probable for $\pi_1 = \pi_2 = .40$ and $\pi_3 = .50$, but Model 2 becomes

the most probable if $\pi_1 = \pi_2 = .40$ and $\pi_3 = .20$. Crucially, there are many reasonable combinations of values that make Model 3 (i.e., the most complex model) more likely to be true than both Model 1 and Model 2. For example, consider the case in which $\pi_1 = .60$ and $\pi_2 = \pi_3 = .70$. Then Model 3 is the most probable (29%), followed by Model 2 (13%) and Model 1 (8%).

In sum: when complexity is evaluated with reasonable criteria, more complex models can be more, less, or equally likely to be true compared with simpler models. Trafimow's argument only applies when informative models that make definite statements about certain variables and causal relations are pitted against less models that exclude those variables from consideration. As I illustrated earlier, this is not helpful if the goal is to figure out the causal structure of the world. The *reductio* that researchers should prefer models involving just two variables and "a single correlation coefficient" brings about additional problems; notably, it becomes impossible to control for the distorting effect of confounders (see Kline, 2016; Pearl et al., 2016).

But even if the complexity of a causal model is not a reliable guide to its plausibility, it may still be the case that one's theoretical model—however simple or complex—has a low probability of being true *a priori*. What should one make, then, of Trafimow's claim that researchers will seldom be able to gather enough evidence to turn an unlikely model into a well-supported one? This assertion rests on a failure to appreciate that likelihood ratios relate to effect size and sample size in a highly nonlinear fashion. In the original paper, Trafimow (2017) seemed to imply that a LR of 50 is often prohibitively hard to obtain: "To take the worst-case scenario illustrated in Fig. 2 […] even an LR of 30 will be insufficient to instill confidence in the model (the required LR is 50!)" (p. 745). Similarly, Saylors and Trafimow (2020) considered a range of LR values from 10 to a maximum of 30, while claiming that even a LR of 10 "can be argued to overstate the quality of most data" (p. 9).

These statements are puzzling because, in practice, it is easy to obtain likelihood ratios higher than 10 or 50—even by several orders of magnitude. To illustrate, the statistical package JASP (JASP Team, 2019) can be used to calculate the marginal likelihood ratio (or *Bayes factor*) for a single correlation coefficient, by pitting the hypothesis that two variables are correlated (with default priors on the size of the effect) against the hypothesis that the correlation is zero. With a sample size of $N = 200$, a sample correlation of $r = .25$ yields a marginal LR of about 50 in favor of the effect. With $N = 200$ and $r = .30$, the ratio increases to about 1,000. With $N = 200$ and $r = .40$ (hardly exceptional evidence), it exceeds *two million*. Note that, using a posterior ratio of 10 as a conventional cutoff, a LR of 1,000 would provide strong support for models with a prior probability of just about 1%. A LR of one million would support models with prior probabilities as low as 0.001%.

Causal models are usually not tested piecemeal (e.g., one correlation at a time), but compared with one another based on their global ability to account for the data. A common approach is to employ information criteria such as the AIC (*Akaike information criterion*) and BIC (*Bayesian information criterion*). If Δ is the absolute difference between the AIC or BIC statistics of two models, $e^{\Delta/2}$ approximates a Bayes factor that compares them (note that AIC and BIC imply different kinds of priors; see Weakliem, 2016). Thus, $\Delta = 5$ corresponds to a marginal LR of about 12 in favor of the best-performing model; $\Delta = 15$ corresponds to a ratio of about 1,800; and $\Delta = 30$ (far from an exceptionally large difference) corresponds to a ratio of

over three million. Of course, convincingly testing alternative causal models of a phenomenon is an arduous task that cannot be reduced to a single statistic, and typically requires the use of multiple methods and study designs (see e.g., Kline, 2016; Pearl et al., 2016; Wiedermann & von Eye, 2016). The point of these examples is that even models with very low probability can receive considerable support, without the need to collect implausible amounts of evidence.

To conclude, Trafimow's argument does not deliver the epistemological leverage promised in the paper. When complexity is evaluated with reasonable criteria, more complex models of a phenomenon can be more, less, or equally likely to be true compared with simpler models. Instead, one is left with the correct but trivial observation that less informative models are more likely to be true than more informative ones. And while interesting theories are often unlikely *a priori*, it can be relatively easy to obtain enough data to support or disconfirm them, without the need for extraordinary evidence. Because Trafimow's argument is meant as a general heuristic in favor of simple causal models, I leave aside the deeper issue of whether researchers should be concerned with the literal truth of a model or merely with its verisimilitude, particularly in the "softer" disciplines (see e.g., Meehl & Waller, 2002).

To be clear, I am *not* saying that theories and modeling practices in the behavioral sciences are just fine as they are. The proliferation of overly complex, poorly justified, and weakly supported models in the literature is definitely a concern (see Saylors & Trafimow, 2020). Parsimony is important, and explanatory theories should not be more complex than they need to be to adequately explain the phenomenon at hand; but the plausibility of any given causal model can only be evaluated on a case-by-case basis. While simple theories are desirable all else being equal, reality is the ultimate arbiter. A survey of successful theories across disciplines would reveal many cases of genuine, intricate complexity; human behavior is hardly going to be an exception. If the ultimate goal is to improve the quality of our science, telling researchers that sophisticated theories "pose a major danger to truth" just because of their complexity (Saylors & Trafimow, 2020) may do more harm than good.

References

JASP Team (2019). *JASP* (Version 0.11). https://jasp-stats.org/

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed). Guilford.

Meehl, P. E., & Waller, N. G. (2002). The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude. *Psychological Methods, 7,* 283-300. https://doi.org/10.1037/1082-989X.7.3.283

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.

Saylors, R., & Trafimow, D. (2020). Why the increasing use of complex causal models is a problem: On the danger sophisticated theoretical narratives pose to truth. *Organizational Research Methods,* 1094428119893452. https://doi.org/10.1177/1094428119893452

Trafimow, D. (2017). The probability of simple versus complex causal models in causal analyses. *Behavior Research Methods, 49,* 739-746. https://doi.org/10.3758/s13428-016-0731-3

Weakliem, D. L. (2016). *Hypothesis testing and model selection in the social sciences*. Guilford.

Wiedermann, W., & Von Eye, A. (2016). *Statistics and causality*. Wiley.