Supporting Information. Grace, J. B., and K. M. Irvine. 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology.

Appendix S2

A Scientist's View of Why Averaging Parameters Across Models Will Produce Muddled Inferences

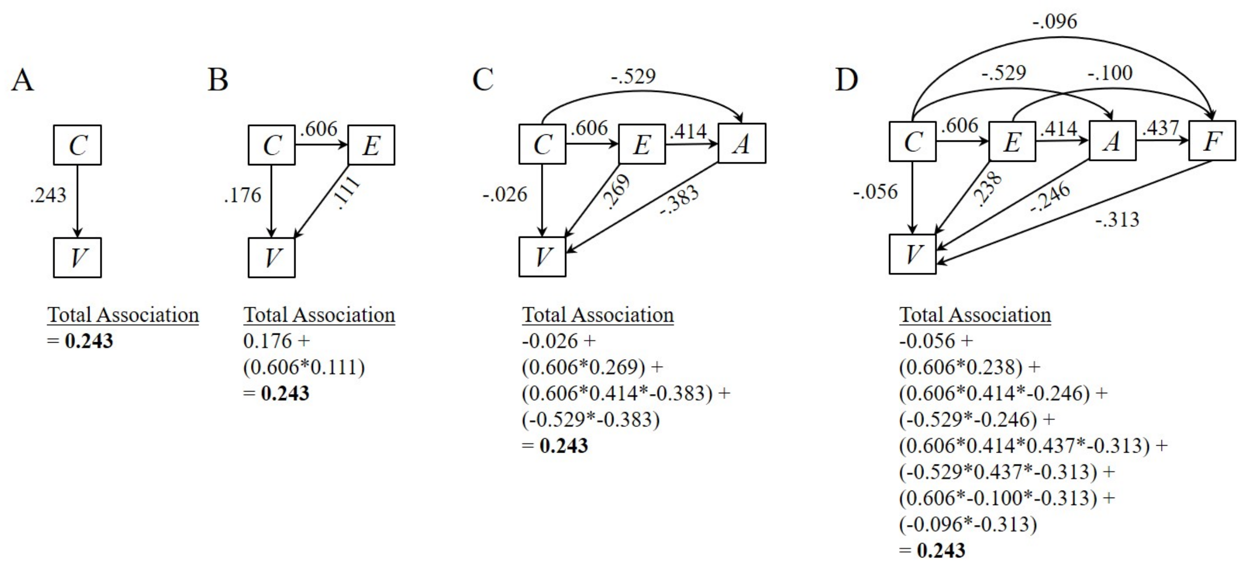
Here we use the hypothesized model structure in Figure 5C of the main text and the wildfire data to show how the rescaling of relationships described by Cade (2015) can be interpreted as the shifting causal meaning of coefficients in competing models. To do this, we use Wright's (1921) method for decomposing correlations and assume Gaussian distributions for both the response and predictor variables.

Figure S1 presents a series of models, with each subsequent model including an additional predictor variable. In this illustration, variables are added to the model in an order that respects the authors' hypothesized causal chain. This of course would not be the typical sequence for variable entry in an automated model-building regression package. In that situation, variables are typically brought in based on their ability to explain remaining variance. Nonetheless, model comparisons often involve comparison among a liberal number of models. In this exercise, we present standardized parameter estimates, which creates a direct comparability with the decomposition of the bivariate/marginal correlations presented in Table 1 of the main text.

Figure S1A represents a model that estimates the net (total) effect of distance from coast C on vegetation recovery V. In this illustration, bringing in additional variables "unpacks" or permits separation of some of the different ways C can influence V. In Figure S1B, the total effect of distance on recovery is partitioned into (a) the indirect effect through increasing elevation with increasing distance from the coast and (b) the remaining effect, which is observed as a 'direct' effect in the model. Comparing Figure S1A with S1B, the meaning of the arrow from C to V changes from the total effect of C on V to the effects of C on V that are independent of the change in elevation expected as E changes with C. We can also see in Figure S1B that the arrow from E to V represents the effect of elevation independent of other influences associated with location. We can quantify the indirect effects by multiplying the coefficients along the path. Thus, in Model B the total effect of C on V in Model A is decomposed into an indirect effect through E, which equals 0.606*0.111, or 0.067, and a residual direct effect of 0.176. Summing the two path strengths returns the total effect, 0.243.

Bringing in the variables stand age (in Model C) and fire severity (in Model D) leads to some remarkably different inferences! Comparing Model C to Model B, we see that the sign of the effect of C on V changes from moderately strong and positive to vanishingly small and negative. This is a clear demonstration of the mysterious "Simpson's Paradox" described in the statistics literature and discussed by Pearl and MacKenzie (2018) at length. Within the interpretive framework built here, we actually expected the partial effect of C on V to be very weak or nonexistent once other conditions were taken into account (refer to the hypothesized effect for path 7 discussed in Table 5). If the reader examines the many differences in coefficients among models in Figure S1, they will notice that some coefficients are unvarying and can be pulled directly out of the correlation matrix (e.g., effect of *C* on *E*). While there is much interesting science here, the point of this exercise is to show how the rescaling of coefficients described by Cade can be understood in this case as the decomposition of various mechanisms relating *V* to *C*.

Figure S1. Set of models differing in the variables included in the model. Results shown here are produced using the data in “Data S1” and the code in “Data S2”. Analyses were run using R version 3.5.3 (R Core Team 2019).



Literature Cited

Cade, B. S. 2015. Model averaging and muddled multimodel inferences. Ecology 96:2370-2382.

Pearl, J., and D. MacKenzie. 2018. The book of why: The new science of cause and effect. Basic Books, New York, NY, USA.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Wright, S. 1921. Correlation and causation. Journal of Agricultural Research 10:557–585.