From: Dennis Yi Tenen
Subject: Memo 001: On Study Design in Computational Humanities
Date: May 10, 2025

CUØ /'kju:nɪl/

Reading Thad Dunning's *Natural Experiments in the Social Sciences* (Cambridge, 2012) I am particularly struck by his discussion of study design. "How can causal inference be improved?" he asks on page 4 and answers: "In seeking to answer such questions, I place central emphasis on natural experiments as a 'design-based' method of research — one in which control over confounding variables comes primarily from research-design choices, rather than *ex post* adjustment using parametric statistical models (4)."

This approach seems particularly well-suited for computational study in the humanities, where "the veracity of causal and statistical assumptions that are often difficult to explicate and defend — let alone validate." The natural experiment approach seeks to shift reasoning about such assumptions from the statistical modeling part of the research process, expressed mathematically, to the design process, expressed in the logic of the world observed: "With natural experiments, it is the research design, rather than the statistical modeling, that compels conviction."

For this reason, Dunning writes, "substantive and contextual knowledge plays an important role at every stage of natural-experimental research — from discovery to analysis to evaluation." The emphasis on context necessitates thinking about statistical concepts such as "effect," in more specified, historical terms. The influence of one author on another, for example, depends crucially on contingent facts about their biography, their publication history, ideology, genre conventions, and numerous other factors worthy of consideration. The design-approach asks us to ground abstract statistical relationships firmly within concrete historical contexts and detailed interpretive frameworks.

As a consequence of reasoning about complicated contexts, the quantitative analysis of natural experiments tends to be simple. Dunning writes: "Often, a minimum of mathematical manipulation is involved. For example, straightforward contrasts between the treatment and control groups — such as the difference in average outcomes in these two groups — often suffices to provide evidence of causal effects (105)." The potential simplicity of quantitative data analysis makes the statistical results easier to convey and interpret, Dunning writes. "Rather than presenting the estimated coefficients from multivariate models in long tables of regression results," he concludes, "analysts may have more space in articles to discuss the research design and substantive import of the results. I would add this also makes them easier to peer-review.

Simplicity ultimately breeds transparency. Again, Dunning: "Analyzing data from strong research designs — including true and natural experiments — requires analysts to invoke assumptions about the process that gives rise to observed data (106)." For me, here finally lies the subtle but crucial point of his argument: all of the above remains true not just for natural experiments, but for strong research study design in computational humanities and social sciences more generally. Christopher H. Achen makes a similar point in his wonderful paper on "Garbage-Can Regressions," arguing for "sophisticated simplicity" in study design, engaging more "creatively" with the data.

The study-design mindset fits well with my organic inclinations as a humanist. I don't normally reason by data manipulation. Reasoning by data manipulation alone risks "cooking the books" in losing sight of the underlying social or linguistic dynamics. The vagrancies of culture force me

to think contextually: in terms of processes, timelines, customs, genres, relationships, narratives, etc. And I would like to remain firmly grounded in that realm when doing computational research.



DENTEN
Director, Narrative Intelligence Lab
Columbia University

Works Cited

- Achen, Christopher H. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22, no. 4 (2005): 327–39. https://doi.org/10.1080/07388940500339167.
- Dunning, Thad. *Natural Experiments in the Social Sciences: A Design-Based Approach.* Cambridge: Cambridge University Press, 2012.

Further Reading

- Gelman, Andrew, and Aki Vehtari. *Active Statistics: Stories, Games, Problems, and Hands-on Demonstrations for Applied Regression and Causal Inference.* Cambridge: Cambridge University Press, 2024.
- Hargittai, Eszter, ed. *Research Confidential: Solutions to Problems Most Social Scientists Pretend They Never Have.* Ann Arbor: University of Michigan Press, 2009.
- Huntington-Klein, Nick. *The Effect: An Introduction to Research Design and Causality.* 2nd ed. Boca Raton: Chapman & Hall/CRC, 2025.
- Knobe, Joshua, and Shaun Nichols. "An Experimental Philosophy Manifesto." In *Experimental Philosophy*, edited by Joshua Knobe and Shaun Nichols. New York: Oxford University Press, 2008.
- Tukey, John W. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley, 1977.