The bigger concern is the lack of control researchers have over the release of the experiments. Groups still fear (legitimately) that ”disappointing” findings will negatively affect their organization. This dynamic could lead to the selective disclosure of experimental results. For illustrative purposes, suppose 20 organizations were testing the effectiveness of direct mail for altering vote choice with researchers. Suppose further that 19 of these experiments came back with null (or even negative) findings and only one showed a positive and statistically significant result, which would be perfectly consistent with a true effect of zero for mail on persuasion. However, if the organization only let the researcher publish the results of the one ”positive” result, readers would be left with the mistaken impression that mail was very effective in persuading voters. The gate keeping authority of the organization is what drives this version of publication bias rather than the researcher's personal file drawer problem (Gerber et al. 2001). The one researcher releasing her results may be completely unaware her findings are spurious, however, the net bias in the published literature is the same.

So what can be done to combat this publication bias? My general practice has been to negotiate the right to publish the results of the experiment up front and promise the organization that their identity will not be revealed. This practice allows me to publish less than flattering results (to the extent that null findings are publishable, see Gerber and Malhotra 2008) but is not a cure all. First, while most readers will not be able to guess the identity of the organization studied, funders will almost assuredly come across the paper and figure out the identity of the organization since they are familiar with the work and researcher / organization networks. Second, organizations are increasingly viewing the results of experiments as valuable and secrets to be kept from competitors. Thus, many simply will not agree to joint ownership of the data and the researcher has to hope the organization will grant release of the proprietary data.

The medical sciences handle the problem of publication bias with proprietary data by requiring all experiments to be registered if they are to be published or used for approval. This solution does not quite work in political science since campaigns do not require FDA approval to implement techniques learned through experiment. Furthermore, registering experiments may tip off competitors on the activities of competitors putting the campaign at a strategic disadvantage. Thus, such a registry would serve as a strong disincentive for campaigns collaborating with academic researchers and curtail the growth of that literature. Perhaps it would be possible to link the experiment to the researcher and make the existence of each experiment public after a period. However, many organizations will not trust that the experiments will not be linked back to the organization, so the chilling effect will only be slightly dampened.

A more modest proposal would be to require researchers to disclose when experiments have a proprietary provenance. A simple statement that the campaign owns the data and allowed the researcher to publish the results will allow the reader to update the probability of publication bias. The magnitude of this bias will be difficult to assess without many publicly available replications, but the researcher may not even know the distribution of treatment effects. Such ownership disclosures may not solve the problem presented by organizations selectively releasing proprietary results, but by alerting readers to the problem it puts experiments conducted by campaigns on the same level of confidence as other experiments.

References

Gerber, Alan S., Donald P. Green and David Nickerson. 2001. “Testing for Publication Bias in Political Science,” Political Analysis 9(4):385-392. Gerber, Alan S. and Neil Malhotra. 2008. “Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals.” Quarterly Journal of Political, Science 3(3):313-326.

Kevin Clarke
Department of Political Science
University of Rochester
kevin.clarke@rochester.edu

## Neyman-Rubin...O'Reilly?

Counterfactuals seem to be very much on the minds of political scientists these days. Whether in the back of dingy hotel conference rooms, the august pages of the *American Political Science Review*, or in faculty meetings amongst the weak coffee and damp danish, there are demands to consider the counterfactual and arguments over what the appropriate counterfactual actually is. The roots of our obsession can be traced to two beliefs that permeate nearly every corner of political science. The first is that Causal Inference should be, and is, the goal of all inquiry in political science. (This belief always puts me in mind of that E*Trade commercial where the voice over intones "No one ever woke up one morning and said 'I just want to be an ordinary scientist,' or 'an ordinary writer,' or even 'an ordinary runner,' " while images of Stephen Hawkings, Ernest Hemingway, and Jackie Joyner Kersee scroll past. I imagine an ad that goes "No political scientist ever woke up one morning and said 'I just want to do description,' " while images of Bill Riker and Warren Miller scroll past.) The second is that the problem of causality has been solved, and the solution involves counterfactuals. There is little that I can do about the first belief in these pages, but there may be something I can do about the second.

My purpose here is not to say that counterfactual accounts of causation are wrong and that some other kind of account is right. Rather, I intend to show that counterfactual analyses are plagued by the same problems that beset all other analyses, namely, they simultaneously admit too little and exclude too much. Philosophers often do this type of work through the use of counterexamples, and I will do the same. For political scientists uncomfortable with this mode of analysis, think of it as assessing the validity of a variable. Any operationalization captures some parts of the concept in question and misses others. At the same time, any operationalization includes parts of other concepts that should have been excluded. A first step in assessing the validity of a measure is to check its face validity; a useful measure of ideology should place Tom Coburn to the right of Chuck Schumer. Analyzing counterfactual accounts of causation works in the same way. We have commonsense notions of what is, and is not, a cause, and valid accounts of causation should be able to distinguish between the two. After all, if counterfactuals do not perform well when we have good intuition, why would we expect them to perform well when we are bereft of intuition and must turn to statistics?

Let us begin with an example in which the counterfactual account admits too little. One of the mantras of those who adopt the counterfactual approach is that there exists "no causation without manipulation." That is, no event is to be considered a cause unless it can be experimentally manipulated. Thus, Holland concludes that attributes of units, such as race or gender, cannot be causes because they could not serve as a treatment in an experiment.[1]

Time for our first counterexample. Most schoolchildren will tell you that tides are caused by the gravitational force of the moon.[2] The moon's gravitational force, however, cannot be manipulated, and it is difficult to see how it could be used as a treatment in an experiment. The counterfactualist must therefore conclude that he does not know what causes the tides, which puts him in league with Bill O'Reilly, who infamously stated that the tides cannot be explained by science. (I have to assume that aligning oneself with the forces of darkness and ignorance is not a comfortable place for most political scientists.)

Astronomy is pretty far removed from political science, but analogous situations are easily found. A large proportion of the rational choice literature concerns the effects of institutions on behavior. Whether of the brick and mortar kind or not, most institutions are not amenable to experimental manipulation. Institutions are generally integral parts of systems in the same way that the gravitation force of the moon is part of the interplanetary system, and for the counterfactualist, attributes of systems cannot be causes. Nonetheless, institutions often play a significant role in shaping and constraining the behavior of actors, and few of us would hesitate to label institutions as causes.

Now let us consider cases where counterfactual accounts admit too much. The following three counterfactuals are political science variations on examples given by Kim (1973).

[1] It is not enough for the attribute to be in principle manipulable. Holland argues that changing an attribute of a unit means that it is no longer the same unit.

[2] Technically, the interaction between the gravitational forces of the moon and the sun and the rotation of the Earth.

1. If German troops had not crossed the German-Polish border, Germany would not have invaded Poland.
2. If we had not held a vote, we would not have passed the budget.
3. If I had not pulled the lever in the voting machine, I would not have voted for Obama.

All three examples are perfectly acceptable counterfactuals, and most counterfactual accounts of causation would admit the premise of each as being causal. In all three examples, however, our intuition tells us that there is no causal relationship. In example (1), the dependency is tautological; troops crossing a border is the definition of an invasion. The real cause of the invasion lies in the strategic situation facing Germany in 1939. In (2), one event, holding a vote, is a constituent part of the second event, passing the budget. Here the cause lies in balancing the interests of the chamber. In (3), an agent, the voter, does an action, voting for Obama, by doing another action, pulling the lever of the voting machine. The cause of voting for a candidate, however, is not in the action of voting, but the real or expected stream of benefits received by the voter. A more useful account of causation would rule (1)-(3) out.

There are still other problems that plague counterfactual accounts. Consider the following:

> If Archduke Ferdinand had not been assassinated, World War I would not have started.

Even if we assume that this counterfactual is true, it requires a number of other counterfactual "causes" to be true as well. The assassination would not have caused the war if the alliance system was not in place, if there had been no arms race, or if nationalism had a weaker hold on the continent. The point is that the context within which the assassination occurred determines, in part, whether the assassination is a cause of the war.

Overdetermination is another problem for counterfactuals, one that proponents have acknowledged. An event is overdetermined if there are two or more sufficient causes of it. Mackie (1974) provides the following illustration:

> Lightning strikes a barn in which straw is stored, and a tramp throws a burning cigarette butt into the straw at the same place and at the same time: the straw catches fire.

The counterfactual "if lightning had not struck, the straw would not have caught fire" is false; the cigarette butt would have started the fire. Similarly, the counterfactual "if the tramp had not thrown the cigarette butt, the straw would not have caught fire" is also false. A counterfactual analysis therefore concludes that neither the lightning nor the cigarette butt is the cause of the fire. Our intuition, however, tells us that both caused the fire. If you think over determination does not occur in political science, consider the causes of voting democratic by an overeducated son of New England, with professors for parents, who attended Oberlin College and the University of Michigan.

To reiterate, my purpose in rehearsing these problematic examples is not to claim that counterfactual accounts of causation should be abandoned in favor of some other account. I want to emphasize that counterfactual accounts are, in some sense, no better or worse than other accounts, and the best known of these accounts continues to evolve in attempts to deal with these and other problems (see Lewis 2000)). All accounts of causation are caught between the Scylla of ruling in events we normally dismiss as causes and the Charybdis of ruling out events we normally think of as causes.[3] There is little reason to treat counterfactual accounts as privileged, and proponents who sniff that counterfactuals provide the only *principled* way of thinking about causation should be dismissed out of hand (see Eells (1991) on probabilistic causation for an alternative). If we cannot privilege counterfactual accounts of causation, we cannot privilege the statistical frameworks, such as the Neyman-Rubin model, that depend on them. So next time someone on the other side of the damp danish demands to know the counterfactual, ask why he or she agrees with Bill O'Reilly about the tides.

[3]The analogy is not perfect; in Greek mythology, Charybdis was considered the greater danger.

Eells, Ellery. 1991. Probabilitistic Causality. New York: Cambridge Univer- sity Press. Kim, Jaegwon. 1973. “Causes and Counterfactuals.” Journal of Philosophy 70 (October): 570-572. Lewis, David. 2000. “Causation.” Journal of Philosophy 97 (April): 182-197. Mackie, J.L. 1974. The Cement of the Universe. New York: Oxford Univer- sity Press.

## Selective Trials

Sylvain Chassang
Princeton
chassang@princeton.edu
Gerard Padro i Miquel
LSE
g.padro@lse.ac.uk
Erik Snowberg
Caltech
snowberg@caltech.edu

### 1 Introduction

Political scientists use randomized controlled trials (RCTs) to address concerns that people who opt into a certain treatment—say, subscribing to a newspaper—have unobserved, systematic differences from those who do not. An RCT randomly assigns treatment, avoiding selection bias.

However, it is difficult to interpret the results of an RCT in the presence of unobserved subject effort. Those who are, say, randomly assigned to receive a newspaper subscription may be unlikely to read the newspaper when it arrives. If an experimenter then observes that those who received a free newspaper subscription have the same level of political knowledge as those who did not, she might wrongly conclude newspapers are ineffective at conveying political knowledge, when it may be that no one in the experiment actually read the newspaper.

In Chassang, Padró i Miguel and Snowberg (forthcoming) we introduce a new framework for experimental design. The designs identified using this framework, which we call selective trials, allow experimental subjects to express preferences over treatment—to control for unobserved effort—while maintaining randomization—to control for selection. Thus, our approach creates designs for analyzing heterogeneous treatment effects, rather than ex-post statistical tools. This article introduces the basic concepts of selective trials through a simple example, briefly discusses what is possible in more general environments, and concludes with some simple guidelines about how to determine whether selective trials may be useful in an experiment.

### 2 A Simple Example

To understand how selective trials can improve inference, we must first formalize the example above of an experiment to evaluate whether newspapers increase political knowledge. This can be seen as a highly stylized version of the experiment in the excellent study by Gerber et al. (2009), in which we consider only a single newspaper—rather than two—and a single outcome measure—political knowledge.

Gerber et al. (2009) finds that giving a free newspaper subscription to someone who does not already subscribe to a paper does not change political knowledge. This may be because most people who do not subscribe to a newspaper will not read a newspaper, even if it is given to them for free. However, there may still be some people in this group who value newspaper subscriptions, just less than their price, and would read the newspaper if they