

# Natural Experiments and Pluralism in Political Science

Philosophy of the Social Sciences  
2015, Vol. 45(4-5) 424–441  
© The Author(s) 2015  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0048393115580266  
pos.sagepub.com



**Sharon Crasnow<sup>1</sup>**

## **Abstract**

Natural experiments are an increasingly popular research design in political science. This popularity raises a number of questions. First, what are natural experiments and why are they appealing? Second, what makes a good natural experiment? And finally, are natural experiments able to provide resources for knowledge production that other methodologies cannot or do not provide? Using Mary Morgan's and Thad Dunning's recent work on natural experiments, I offer answers to the first two questions and use the analysis to argue that natural experiments highlight features of knowledge production that support methodological pluralism and the multiple aims of research.

## **Keywords**

natural experiment, pluralism, methodology, political science, research design

## **I. Introduction**

As Thad Dunning (2012, 1) notes in *Natural Experiments in the Social Sciences*, "Natural experiments are suddenly everywhere." The claim is

---

Received 13 March 2015

<sup>1</sup>Norco College, Norco, CA, USA

### **Corresponding Author:**

Sharon Crasnow, Department of Arts, Humanities, and World Languages, Norco College, Norco, CA 92860, USA.

Email: sharon.crasnow@norcollege.edu

ambiguous. One clear meaning is illustrated by the graph in his book showing the increase in the use of natural experiments in both economics and political science. Political science follows the lead of economics in this trend as it so often follows the methodological lead of economics. But there is another meaning of Dunning's claim. Political scientists are *seeing* natural experiments everywhere—sometimes even when it is a stretch to call the research design a “natural experiment,” in part because of their popularity. Considering this second reading, we may well ask when it is legitimate to make that stretch from the natural site to the natural experiment research design and what political scientists might be gaining by doing so. To answer these questions, I will first address three others. What are natural experiments and why are they appealing? Second, what makes a good natural experiment? And third, are natural experiments able to provide resources for knowledge production that other methodologies cannot or do not provide as well?

I begin by offering an answer to the first question; it is actually a complex question, in that it asks both for an account of what natural experiments are and an explanation of their appeal. I use Mary Morgan's (2013) distinction between Nature's or Society's experiments and social scientists' natural experiments as a starting point for understanding some ways we might think of natural experiments. Dunning's analysis of natural experiments, when seen through Morgan's distinction, suggests that we think of natural experiments along a continuum with better and worse arguments made for their ability to provide evidence in support of causal claims. The degree to which they are able to do so rests in their similarity to true experiments, particularly to randomized controlled experiments or randomized controlled trials (RCTs).

To interrogate this comparison of natural experiments with RCTs, I consider the criteria along which arguments for the evidential effectiveness of natural experiments can be evaluated. In doing so, we see that this evaluation also requires transparency of research design, which in turn depends on case knowledge of the site treated as a natural experiment. The requirement for transparency thus proves to be a virtue of (good) natural experiments. The focus on supporting and testing causal claims, causal mechanisms, and causal principles—general claims or principles—in political science research has often meant that contextual aspects of knowledge production and use are downplayed or ignored. Considering natural experiments and their design highlights ways that a diversity of goals may be served through sensitivity to context. Thus, I conclude that understanding natural experiments helps to reveal a connection between methodological pluralism and pluralism about the aims of science. Different methods are able to reveal different features of the world that are relevant to experimental design, what counts as evidence,

and consequently what conclusions can be drawn from the evidence produced by natural experiments. Furthermore, the insights that come from looking at natural experiment research design have a bearing on other social science research methods.

## **2. What Is a Natural Experiment?**

Shifts in methodology within a field typically are driven by a belief that the new or trending methodology is better at achieving the goals of the discipline. Indeed this is what has driven the turn to experiments in political science more generally, including natural experiments. There are several, related reasons for the rise of experiments in political science. First, political scientists appear to be committed to the idea that hypothesizing, testing, and perhaps even establishing general causal claims is the primary aim of research in their discipline. Second, observational methods—both traditional case studies research and statistical methods—have each been challenged over their ability to achieve this goal. Case study research has been criticized in that supporting causal claims from one or even several cases is problematic. Statistical methods (particularly multivariate regression analysis, a preferred statistical technique for investigating the relationship between more than one independent variable and some dependent) have been increasingly scrutinized as the assumptions that they require have been challenged (Freedman 2010). The move to experiments holds out the promise of escaping difficulties associated with both qualitative and quantitative (primarily statistical) methods, a virtue noted by Morton and Williams (2010) in their guide to experiments in political science, targeted at social science researchers.

But until fairly recently, field experiments have been thought to be difficult or impossible to conduct in the social sciences, and both field experiments and lab experiments may raise ethical concerns. Natural experiments hold out the promise of avoiding some ethical worries given that the intervention does not take place for the purpose of studying the effects—“nature” conducts the experiment rather than the researcher. The events have typically already occurred and so worries about initiating harms are not relevant. Worries about the problem of the small  $n$ —samples that would typically be thought too small to support inference—are also easier to circumvent through seeking natural experiments that offer a larger sample size. Dunning (2012) has argued that natural experiments may also escape some of the problematic assumptions made in the use of multivariate regression analysis, because it is often possible to use a simple difference-in-means test—a test to determine whether the mean measured effect in the experimental group differs significantly from that of the control

group—thus potentially avoiding the more problematic assumptions of more sophisticated statistical techniques.

What is a natural experiment? Natural experiments are not true experiments but rather observational studies.<sup>1</sup> They get their honorific title from their methodological and epistemological proximity to true experiments, most particularly to randomized controlled experiments (or RCTs), particularly those that are field experiments. Natural experiments parallel RCTs, in that there are two groups—a treatment and control group—and an intervention that the treatment group is exposed to and the control group is not. Natural experiments differ in that (a) the groups are distinguished by “naturally” occurring circumstances,<sup>2</sup> (b) the assignment to the two groups is random (as it is in RCTs) or “as if” random (not the case in RCTs), and consequently, (c) the intervention is not in the control of the experimenter. As political scientists Sekhon and Titiunik (2012, 53) put it,

Natural experiments are situations in which an intervention is—in the best-case scenario—randomly assigned, but this intervention is not under the control of the researcher. Rather, it is assigned by nature or by individuals whose goals differ from those of the researcher.

That neither the intervention nor the assignment is in the control of the researcher are thus the key differences between natural experiments and true experiments.

One way to think about how natural experiments are discovered and designed is to use the distinction that Morgan (2013) makes between natural experiments and Nature’s or Society’s experiments. The latter are events that occur in nature or society that create circumstances that might be seen as resembling the experimental setting and thus may provide a site for the investigation of causal complexes. The former are more deliberate reconstructions of these events or circumstances to test or provide support for a particular

---

<sup>1</sup>Dunning (2012, 16) claims that the key element of true experiments that natural experiments lack is that there is no experimental manipulation of the treatment or the research groups.

<sup>2</sup>“Naturally occurring” is ambiguous for the social sciences, because the sites of natural experiments are generally in society and the result of conscious policy decisions. Natural experiments should be distinguished from field experiments where policy decisions are made in conjunction with researchers seeking to find out the effects of a particular policy. Natural experiments are retrospective—researchers discover the sites after the intervention has already occurred and conceptualize the events in terms of an experiment.

causal claim. Although my terminology up till this point has been loose, it is Morgan's more deliberate reconstructions by social scientists that I have been referring to as "natural experiments."

Nature's or Society's experiments arise in a variety of circumstances as detailed by Morgan. It is the way in which these circumstances appear to meet *ceteris paribus* conditions that makes it plausible to think of them as experiments. Following Marcel Boumans (1999), Morgan (2013) distinguishes (a) *ceteris neglectis*, in which the intervention is so great as to overwhelm all other potential causes; (b) *ceteris absentis*, in which the all other potential causes are missing because of the isolation of the experimental group; and (c) *ceteris paribus*, all other potential causal factors are the same in the two groups and so they differ only in terms of the intervention/treatment. Four circumstances that offer Nature's/Society's experiments are massive interventions that make issues of control irrelevant, situations of total isolation, unusual events that occur within a stable environment, or situations in which both the environment and the intervention are carefully controlled through natural or social circumstances (Morgan 2013, 345-46). A massive intervention produces a "before" and "after," with the intervention being so powerful as to swamp all other factors (*ceteris neglectis*). An earthquake might provide this kind of opportunity. Islands may produce situations of total isolation (*ceteris absentis*) and so may provide the opportunity for comparison with a similar island or islands in which some differing factor (or factors) are present. Darwin's use of the Galapagos Islands in this way is a well-known example. An unusual event in a stable environment differs from a massive intervention in that the argument is not that other factors are irrelevant but that they remain the same (*ceteris paribus*). Fisman (2001) uses Suharto's death to produce evidence that political connectedness is causally relevant to the value of firms (stock prices). He argues that "[b]ecause of Indonesia's highly centralized and stable political structure (until the very end of Suharto's reign), it is possible to construct a credible index of political connectedness" (Fisman 2001, 1095). And finally, social circumstances where all other factors are controlled for reasons other than those relevant to the effect of the intervention present the opportunity for a natural experiment as well.

Morgan uses Robert Merton's investigation of mass propaganda through examining the one-time extraordinary jump in sales of war bonds after a radio-marathon drive featuring the singer Kate Smith in 1943 as an example of a Nature's or Society's experiment. As Morgan (2013, 349) notes, "[p]erhaps the best way to describe Merton's investigations is that he, as scientist, dug into the gardens of this Society's experiment to understand how the whole process worked, not just one hypothesis about one bit . . ." In a

Nature's/Society's experiment, some event or circumstances provide the opportunity for an investigation—a case through which we might uncover some understanding of how a social phenomenon works within its “natural” setting, which includes complex interrelations. However, it does not provide a site for testing a particular causal hypothesis without a reconceptualization of the circumstance or data. To turn a Nature's/Society's experiment into a natural experiment, the circumstances have to be re-described or reconstructed in such a way that they can be shown to be closely analogous to an experimental design. The language used in discussing natural experiments reflects this. When a researcher construes a particular site of research as a natural experiment, the research design is often described as “clever.” There are two types of cleverness involved here. The first is in the original identification of the site. Not everyone is able to see that the circumstances could be used as a site of investigation. The second sort of cleverness comes with the reconceptualization of the Nature's/Society's experiment into a natural experiment—the research design.

A Nature/Society's experiment can be investigated “as is”—as a complex natural or social phenomenon. Morgan describes Merton's use of the bond drive in this way. In such a case, the exploration of the event involves little or no reconceptualization beyond conceiving of it a research site. But when such events become natural experiments as the researcher seeks to isolate a causal relation, arguments to support that and to what extent the methodology resembles an experiment are needed.

Experiments are thought to have the power to isolate causal relations through the control of the circumstances in which they are conducted and the manipulation of the supposed causal factor. Arguing that something similar has been accomplished in a natural experiment is arguing to the method's effectiveness for supporting a causal claim. Thus, the evaluation of the natural experiment—a determination of how good an epistemic tool it is—is a function of how strong the argument for the analogy to a true experiment is. To evaluate an analogy requires identifying the similarities and difference in the circumstances and thus arguing that a natural experiment is analogous to a true experiment requires making the argument that at least these two conditions apply: (a) there is an intervention—or something that can be thought of as an intervention—that divides the objects of study into two groups; (b) the division is random or as-if random—in other words, no criteria have been applied to make the division that might have an effect on the outcome of interest.

A nice feature of Morgan's distinction is that it reminds us that a natural experiment involves identifying a circumstance that can be conceived of through this research design and that to do so is to find features of the

Nature's/Society's experiment that support an argument that we can so construe the event in this way. Dunning's account of how to discover and evaluate a natural experiment might be thought of as doing just that. He describes finding pieces of knowledge at the site—the Nature's/Society's experiment—as finding “Causal Process Observations” (CPOs).<sup>3</sup> These are nuggets of information, both quantitative and qualitative, that a researcher picks out through knowing the details of the case. That such information can be either quantitative or qualitative suggests that natural experiments also encourage mixed method research (methodological pluralism). These pieces of information, or as Dunning calls them “nuggets,” are information relevant for causal reasoning. A way of translating this idea to our current discussion using Morgan's terminology is that this is information that emerges from an understanding of the specifics of the context of the Nature's/Society's experiment and allows the researcher to construct that site as a social scientist's natural experiment.

### **3. Natural Experiments as Evidence for Causal Claims**

If the best evidence for causality rests in RCTs, then the strength of natural experiments depends on their similarity to that design. Thus, an argument that supports the analogy between a natural experiment and a true experiment is also an argument that the natural experiment is evidence for the causal inference in question. Both Morgan and Dunning give accounts that resemble this way of thinking about how a natural experiment is designed. Morgan (2013, 55) highlights the parallel between true experiments and natural experiments in the following way:

Perhaps the best test of whether a Society's experiment could be turned into a natural experiment lies in whether a field or laboratory experimental design could be conceived to measure and test the independent effect of any single element under investigation.

Dunning takes the evaluation of the natural experiment to be addressed across three dimensions: (a) the claim that the assignment to groups is random or as-if random, (b) the credibility of the statistical and causal models, and (c)

---

<sup>3</sup>Causal process observations (CPOs) appear in the political science literature under that name in Brady and Collier 2004 (Brady and Collier 2010). The idea seems to originate with Alexander George's work on causal process tracing—George and McKeown (1985) and George and Bennett (2005).

the relevance of the intervention to the causal claim. The inferential strength of the natural experiment thus rests on the strength of the arguments that the natural experiment does well across these dimensions. It is important to note that Dunning recognizes that an argument that does well on one dimension may not do well on others. The ideal is to maximize all dimensions, but in reality there are frequently trade-offs across the dimensions. Thus, one important feature of this analysis is that even natural experiments that fall short of the ideal may still have some inferential value, in that they are strong on at least one of the dimensions.

In addition, because of the three dimensions across which the argument must be made, even when a natural experiment does not clinch or even provide strong support for a causal claim, it may still suggest alternative avenues of research, alternative hypotheses, connections to other research done in the field, or otherwise contribute to knowledge production. That is, information from the context used to make arguments across these dimensions may also serve other functions in research. This is, of course, true of other research designs, but it is particularly clear in the case of natural experiments because of the need to make reference to details of context to establish the viability of the site as a natural experiment. The aim of supporting a causal claim is only one among a plurality of goals of research. We can see this in considering how natural experiments work through an examination of the roles that Dunning describes for CPOs.

Dunning (2012, 209) identifies five types of CPOs, each distinguished through the role it plays in the earlier categorization that Mahoney (2010, 127) used to discuss how CPOs are used in testing hypotheses. Dunning's interests extend beyond testing, however, as suggested above. I have adapted his account and also added a sixth role that supports the discovery of the original site (discovery CPOs).

- *Treatment-assignment CPOs*: information that allows the researcher to determine how to think about assignment to the control or treatment group, support for the claim of as-if random assignment
- *Independent-variable CPOs*: information about the presence of a treatment (intervention) (from Mahoney)
- *Mechanism CPOs*: information about whether an intervening event in the causal process is present and about the kinds of causal processes involved (from Mahoney)
- *Auxiliary-outcome CPOs*: information about outcomes other than the main outcome of interest that would also be expected given the causal process examined in the natural experiment (from Mahoney)



- *Model-validation CPOs*: information that provides support for the credibility of the statistical and causal models used
- *Discovery CPOs*: information about the site that suggests that it is suitable to be treated as a natural experiment (my addition)

This categorization is neither meant to be exhaustive nor mutually exclusive. For example, an independent-variable CPO might also be a mechanism CPO, as the independent variable might be a feature of the mechanism. However, these could also be distinct as the identifying features of the mechanism through which an independent variable is connected to the dependent variable will probably include additional elements. In addition, the means for identifying these particular CPOs might vary. It could also be the case that the same information serves more than one function—a discovery CPO may work to alert a researcher to the possibility that the site could serve as a natural experiment, but also serves as an independent-variable CPO when that researcher refines the argument. The following example illustrates both how CPOs work as evidence in an argument that the Nature's/Society's experiment meets the requirements of a natural experiment and its evaluation across Dunning's three dimensions.

#### 4. Hyde: Evidence from a Natural Experiment

In her 2007 *World Politics* article, “The Observer Effect in International Politics: Evidence from a Natural Experiment,” Susan D. Hyde offers a natural experiment as evidence that international election observers had an effect on voting fraud in Armenia in the 2003 election. While the case of this particular election in Armenia serves as the site of her analysis, she situates the value of natural experiment research design in the broader literature. In attempting to establish the effects of international election observers on voter fraud, cross-country (macro) comparisons suffer from an insoluble endogeneity problem—how to separate the effects of differences among countries from the effects of international observers. She proposes a within-country (micro) analysis as a way of investigating the effect of international election observers on voter fraud—in this way, the comparison is presumably like-to-like because the country is compared with itself at two different times. She chooses post-Soviet Armenia, a case that she knows well and argues that it provides a particularly good venue for such an investigation given the relative homogeneity of the country (very little ethnic diversity) and the stability of the issues (the only major political issue was the question of how to treat Azerbaijan). Her choice of this site is thus driven by “nuggets” of information

(CPOs) that she has in virtue of her familiarity with the region, and we see how her familiarity with the specifics of the context functions in her argument at several points. We might think of these nuggets as information that supports the *discovery* of the Nature's/Society's experiment—the information that enables the researcher to identify a site as having some set of properties that would make it suitable for consideration as a natural experiment.

As Hyde recounts, the 2003 elections were widely considered a turning point in Armenian democracy—a point at which voter fraud was seriously addressed. A few months prior to the elections the Armenian Ministry of Foreign Affairs invited the Organization for Security and Co-operation in Europe/Office for Democratic Institutions and Human Rights (OSCE/ODIHR) to send international observers for the first time. This information suggests that the presence of observers might provide an opportunity for investigating causality through a natural experiment (*discovery CPO*, but also *mechanism CPO*—see the *Credibility of Causal Models* below). This is an example of an unusual event occurring against a stable background—one in which election fraud has been the norm. Notice that while knowledge of the introduction of observers as a change from business as usual is a *discovery CPO*, the presence of the observers will also function as an independent variable in Hyde's research design.

#### 4.1. Randomization

Other features of the Armenian case also make it suitable to treat as a natural experiment. Dunning takes as a hallmark of a natural experiment that the assignment to the control and treatment groups is as-if random (an argument made using *treatment-assignment CPOs*). Hyde establishes that the observers, while not strictly randomly assigned, were assigned in a way that approximates randomness (“as-if” random). She investigates the site using both qualitative and quantitative evidence to support that claim—thus she is using mixed methods research (methodological pluralism). For example, as part of her qualitative evidence, she conducts interviews with members of the OSCE/ODIHR through which she determines the criteria used for assignment of observers and other relevant factors. She also conducts randomization checks, first comparing the observed polling stations relative to population, arguing that the distribution of observers was widespread throughout the population, relatively evenly distributed per population in urban and non-urban areas and relative to other potential confounders and so gives no indication of being correlated with a bias against the incumbent using both quantitative (to establish the proportional distribution) and qualitative (to determine the whereabouts of the observers) data. The evidence here also depends on *independent-variable*

CPOs given that information is also being collected on the presence or absence of observers.

In addition, she notes, “Ideally, in any experimental research design the assignment of the treatment could be examined in relation to a background covariate to test for balance between the treatment and control groups” (Hyde 2007, 56). These data were not available for the first round of the election, but as the election required a second round run-off she was able to obtain it for that round:

As a further examination of the “as-if” randomization, round 2 treatment of observers is compared with vote share and turnout in round 1. These round 1 outcomes should be equal between polling stations that were monitored in the second round and those that were not. The results . . . support the assertion that the method used by the OSCE/ODIHR to assign observers approximates randomization. (Hyde 2007, 57)

#### 4.2. Credibility of the Statistical/Causal Models

An important feature of Hyde’s argument is what she takes as a measure of voter fraud. International observers are taken to be causally relevant to voting for the incumbent and so that vote is an indicator of voter fraud (*independent-variable CPO*, but also *mechanism CPO* because the observers are causally relevant within the particular model that includes other factors). Her argument for relevance of the intervention to voter fraud is based on her (qualitative) contextual knowledge of Armenia during the post-Soviet period (*mechanism CPOs*). Only the executive office was organized enough and held the requisite power to perpetrate election fraud. Consequently, she reasons that a significant difference in votes for the incumbent in the control group versus the experimental group is an indicator of fraud—where there is less fraud the percentage of votes for the president will be lower than it will be in polling places where there is substantial fraud.

The credibility of the causal model, that is, the causal connection between the executive power (the sitting president), the control of the polling places, and voter fraud, is argued for based on Hyde’s regional, primarily qualitative, knowledge of post-Soviet Armenia. The credibility of the statistical model is high given that she uses a simple means-difference test—there is nothing statistically “tricky” in the method. This particular election provided an additional opportunity for testing the effectiveness of international observers given that it involved two rounds—an initial election in which none of the candidates received 50% or more of the votes and a run-off. Hyde’s comparison included the control and experimental groups as described from the first election and then comparisons of the same polling stations between the first

and second election. She does find significant differences, determined by difference-in-means through  $t$  tests—a standard significance test for hypotheses. In the first election, she finds a 5.9% difference between the experimental group's and the control group's votes for the president, and in the second election, she finds a 2% difference, both of which are significant and support the finding that international observers lessened election fraud in the 2003 Armenian election (*model validation CPOs*).

Hyde does not “adjust” her groups—that is, she does not reconstruct them using statistical methods. If she had done so, further argument would be needed to support the credibility of the statistical model. She does check the groups that are created through the intervention (the presence of observers) for balance. In other words, she checks them for other possible independent variables that might play a role in generating the dependent variable. The checks for balance involve quantitative analysis (she examines population distribution relative to the distribution of observers); however, to do this she also uses qualitative methods to gather information (interviews with members of the OSCE/ODIHR) to determine both how the distribution was originally accomplished and how observers operated during the election.

In addition to the randomization check described above, Hyde also explores alternative hypotheses that could have resulted in the differences between the two groups. The plausibility of these hypotheses (difficulties that observers might have had reaching polling stations, differences in observations between rural and urban stations, differences in different regions) is also checked both quantitatively and qualitatively and each is rejected after careful consideration (*auxiliary hypothesis CPOs*).

This analysis of the use of CPOs offers one way of seeing how Hyde uses the information gathered through multiple methods to engage in multiple aspects of knowledge production. Because this research design is a natural experiment with the goal of supporting a causal claim—at least about *this* case—that is the focus. But in her argument, she generates information that serves the other goals of knowledge production as well. Her research suggests alternative hypotheses (some of which she tests), reveals features that suggest other applications (ways in which observers might be more effectively deployed, for example), and makes clear factors that ought to be considered if we are to export this knowledge elsewhere. This final point brings us to Dunning's third dimension—relevance.

### 4.3. Relevance

Dunning's third dimension for evaluation is substantive relevance of the intervention. As he puts it, “to what extent do the effects of random or as-if

random intervention in fact shed light on wider social-scientific, substantive, and/or policy concerns?" This final dimension differs from the previous two, in that while they are concerned with fundamental epistemological issues that have to do with the strength of inference, relevance is judged relative to the pragmatic issue of the extent to which the intervention provides evidence for the causal relation of interest. That is, not just what happens in this case, but the more general question of the relevance of international observers to election fraud. Hyde (2007, 63) is cautious about her claims of relevance.

This article does not hypothesize about whether the results can be extended outside of the context of the experiment . . . The results presented here represent one step toward an improved understanding of how international actors influence domestic politics and provide a test of the effectiveness of one of the most widespread democracy-promotion activities carried out by international organizations and international NGOs.

Thus, the broader relevance of the claim that Hyde supports is not particularly strong if we just consider the natural experiment on its own. However, seen in the context of other research on international observers it may become part of a more compelling argument for the effectiveness of international observers. Here we see a trade-off among dimensions. Hyde has made compelling arguments for randomization and the credibility of the causal model. These arguments relied heavily on the details of this particular case and consequently she is cautious in her conclusions about the broader relevance of her findings. As Dunning notes, such trade-offs among the three dimensions along which natural experiments can be evaluated are likely.

## **5. What Methodological Virtues Do Natural Experiments Have?**

The previous sections have been devoted to answering the first two questions—what are natural experiments and what makes a good natural experiment. Some of the elements that have been detailed in the example, together with the more general idea that arguments to the effect that we have a site that can serve as a natural experiment and that aid in its evaluation, suggest other ways natural experiments may contribute to knowledge production. For example, the arguments that natural experiments need to establish their strength also provide information that generates alternative hypotheses, suggests integration with other knowledge projects, is relevant to the transportability of results (the limits of applicability), and so may ultimately support questions of policy effectiveness. These aims of research sometimes receive

less attention than the testing of causal hypotheses, particularly in political science, which takes testing as a primary indicator of the “scientific” nature of the discipline. However, the importance of such reasoning to knowledge production should be recognized and the ways that various research methods contribute should be accounted for. Natural experiments are one venue in which the need to consciously address both of these concerns is made clear. I have argued that because natural experiments are not true experiments, if they are to generate evidence relevant to causal conclusions, they need to appeal to additional arguments. These arguments might be thought of as arguments that the site of a Nature’s/Society’s experiment can be reconstructed or designed as a natural experiment. The need to make the argument that the original site can be understood as a natural experiment under some reconstruction requires knowledge of context in ways that contribute to other aims of knowledge production. The element of randomization or as-if randomized selection of the control and experimental groups combined with the need for detailed knowledge of the context speaks to the value of mixed methods research.

But in fact these are not features unique to natural experiments. All uses of method, including true experiments, and particularly RCTs, require auxiliary arguments that the method is adequate to the phenomena and the goals of explanation. In the case of RCTs, the arguments are thought to have been made through the research design. Because natural experiments highlight the need for supporting auxiliary arguments, they show the importance of mixed methods research to generate and defend the assumptions that causal inference from any method requires. Their suspect nature forces transparency of method.

Morgan and Dunning both make the point that the natural experiment—or the reconstruction of the natural experiment from the Nature’s/Society’s experiment—is only as good as the argument that the conditions that make the circumstances analogous to an experiment are met.<sup>4</sup> But Morgan’s account highlights a caveat—there are very likely Nature’s/Society’s experiments

---

<sup>4</sup>Dunning focuses on random selection as the key criterion of natural experiments and so rejects post-experimental statistical manipulation of the control and experimental groups as part of the reconstruction. Morgan has a more liberal view in which reconstruction of Nature’s/Society’s experiments into natural experiments all involve some manipulation and post-experimental statistical manipulation is as good as the argument that can be made for it. Post-experimental statistical techniques are prevalent in economics, and Dunning’s base in political science may account for some of this difference, although there are also political scientists who take Dunning’s view to be too restrictive (Robinson, McNulty, and Krasno 2009).

where the isolation of a cause cannot be accomplished because of causal complexity or because not enough is known about the case to allow for its isolation. Although such sites may still be fruitful for a more holistic investigation, to treat them as social scientists' natural experiments would be a mistake, according to Morgan. Dunning's account accommodates this point as well because a researcher may not be able to make a good enough argument across one or more of the dimensions of evaluation. In this sense, it is both a virtue and a vice that natural experiments come out of a "natural" setting. In such a setting, there are at least some of the ingredients for producing the evidence that allows the researcher to engage in causal inference and so the complexity provides a rich source of information. But the downside is that the complexity sometimes will not yield to analysis where interconnections between events are difficult to identify or eliminate to hone in on a particular causal pathway.

Another potential virtue of natural experiments is in the way that they both focus on and provide information that is relevant to how well results can travel from the original site to others. True experiments are often criticized because of issues of external validity. What works in the lab or in a field experiment in one place may not work elsewhere. It has been argued that natural experiments are less likely to suffer from issues of external validity, although this seems questionable. As noted in Hyde's Armenian election example, the question of how far the result of the natural experiment travels is up for grabs. But perhaps natural experiments may have an advantage on this front. The details of the circumstances under which the causal relation holds are often present in the arguments that natural experiment design requires and thus the extent to which other circumstances resemble those of the natural experiment can be more readily assessed.

This idea also provides a suggestion about how natural experiments may be particularly good candidates for providing evidence for use, which, as Nancy Cartwright has argued, is not always the same as evidence for belief (Cartwright and Hardie 2012; Cartwright and Stegenga 2011). In the case of policy decisions, the question would be whether support for a causal claim can be used in informing a policy decision—the natural experiment establishes that it works somewhere, but will it work here? There are a variety of considerations that are relevant to a full theory of evidence for use; however, the cornerstone of Cartwright's account is the recognition that in isolating a cause, we are identifying an *insufficient* but *necessary* part of an *unnecessary* but *sufficient* condition for producing the effect. In other words, she treats causes as INUS conditions (Mackie 1965). Thus, the cause is part of a larger causal complex that includes what Cartwright refers to as supporting conditions—the other elements of the unnecessary but (jointly) sufficient conditions that bring about

the effect. Roughly, Cartwright's argument theory of evidence for use requires minimally that there is good evidence to warrant that the particular causal relation holds and that there is evidence that the supporting factors that the cause requires to operate as a cause are in place.

Although Cartwright's primary interest is in the transportability of causal claims that are warranted by RCTs, the idea works with natural experiments as well. In my use of Dunning's criteria, I have been concerned to understand how it is that the causal claims supported by natural experiments are warranted (evidence for belief). The answer to that question has focused on the need for pieces of evidence that address the three dimensions of evaluation, but because natural experiments arise out of natural/social circumstances—they are reconceptualized from Nature's or Society's experiments—the evidence that is needed to establish them as evidence for the causal claim is also often the evidence that is needed to aid in determining whether that causal claim should guide a particular policy decision. That is, it is evidence that the supporting factors are present. Hyde's reluctance to make the general claim that her natural experiment shows the effectiveness of international election observers is appropriate, but the arguments that she presents to support that she has indeed designed a (relatively) strong natural experiment include identification of the support factors that have a bearing on the causal relation. In deciding whether the policy will work elsewhere, at least one thing that has to happen is considering whether the same (or relevantly similar) support factors are in place.

Just as with true experiments (both lab and field experiments), the experiment (or natural experiment) is just as good as the evidential arguments—the arguments that the circumstances are controlled (the randomization claim is true), that the statistical and causal models are credible, and that the intervention is relevant.

## 6. Conclusion

Dunning argues that the dimensions across which we evaluate natural experiments are the dimensions across which all research designs are to be evaluated—both observational and experimental. Hence he argues for the transparency of the circumstances of research, the assumptions that the researcher makes, and an understanding of the methods used in all research to support the extent to which these dimensions have been addressed. The process of probing Nature's/Society's experiments to make the arguments that the site can serve as a natural experiment calls for a detailed examination of the context. This sensitivity to context may have a variety of positive effects, illustrated here in relation to natural experiments, but important



more broadly for the use of all methods. Even inadequate arguments for establishing the site as a natural experiment hold promise for suggesting alternative hypotheses, policies, and the interconnections that give rise to a rich theoretical treatment of phenomenon, such as Morgan suggests the original site might provide. Consequently, even failed natural experiments might be of use in knowledge production. But this too is true of other methods. Hypothesis generation, concept formation, integration of knowledge, and evidence for use all come out of such probing. The current interest in natural experiments in political science thus not only provides an argument for the value of mixed methods (methodological pluralism) in the field but also is a research design that can be a resource for the plural aims of science.

### Acknowledgment

The author thanks Mary Morgan, the participants of the European Network for Philosophy of Social Science Conference, Madrid, September 2014, and the referees of this journal for their comments.

### 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.

### References

- Boumans, Marcel. 1999. "Representation and Stability in Testing and Measuring Rational Expectations." *Journal of Economic Methodology* 6:381-401.
- Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham: Rowman & Littlefield.
- Brady, Henry E., and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. 2nd ed. Lanham: Rowman & Littlefield.
- Cartwright, Nancy D., and Jeremy Hardie. 2012. *Evidence Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press.
- Cartwright, Nancy D., and Jacob Stegenga. 2011. "A Theory of Evidence for Evidence-Based Policy." In *Evidence, Inference and Enquiry*. Proceedings of the British Academy, edited by P. Dawid, W. Twining, and M. Vasilaki, 291-322. New York: Oxford University Press.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design Based Approach*. New York: Cambridge University Press.

- Fisman, Raymond. 2001. "Estimating the Value of Political Connection." *The American Economic Review* 91 (4): 1095-102.
- Freedman, David A. 2010. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Edited by David Collier, Jasjeet Sekhon, Philip B. Stark. New York: Cambridge University Press.
- George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development*. Cambridge: MIT Press.
- George, Alexander L., and Timothy J. McKeown. 1985. "Case Studies and Theories of Organization Decision Making." *Advances in Information Processing in Organizations* 2:21-58.
- Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60 (1): 37-63.
- Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2:245-64.
- Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62:120-47.
- Morgan, Mary S. 2013. "Nature's Experiments and Natural Experiments in the Social Sciences." *Philosophy of Social Science* 43 (3): 342-57.
- Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge University Press.
- Robinson, Gregory, John E. McNulty, and Jonathan S. Krasno. 2009. "Observing the Counterfactual? The Search for Political Experiments in Nature." *Political Analysis* 17:341-57.
- Sekhon, Jasjeet S., and Rocío Tituniuk. 2012. "When Experiments Are Neither Natural Nor Experiments." *American Political Science Review* 106:35-57.

### Author Biography

**Sharon Crasnow** is a professor of philosophy in the Department of Arts, Humanities, and World Languages at Norco College. Her research is primarily on the relationship between methodology and epistemology with particular attention to objectivity, evidence, and feminist philosophy of science. She is the co-editor with Anita Superson of *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy*.