Studies in History and Philosophy of Science 78 (2019) 98–104

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Types of experiments and causal process tracing: What happened on the Kaibab Plateau in the 1920s

Roberta L. Millstein

Department of Philosophy, University of California, Davis, One Shields Ave, Davis, CA 95616, USA

HIGHLIGHTS

- Causal process tracing is more of a practice than it is a specific method.
- Causal process tracing can be used in conjunction with different experiment types.
- Different cause-finding practices can be combined to increase causal confidence.
- The wolf-deer populations of the Kaibab in the 1920s demonstrate these claims.
- Causal process tracing can be profitably deployed outside the social sciences.

ARTICLE INFO

Keywords:
Experiment
Causal process tracing
Causation
Methodology
Trophic cascade
Ecology

ABSTRACT

In a well-cited book chapter, ecologist Jared Diamond characterizes three main types of experiment performed in community ecology: laboratory experiment, field experiment, and natural experiment. Diamond argues that each form of experiment has strengths and weaknesses, with respect to, for example, realism or the ability to follow a causal trajectory. But does Diamond's typology exhaust the available kinds of cause-finding practices? Some social scientists have characterized something they call "causal process tracing." Is this a fourth type of experiment or something else? I examine Diamond's typology and causal process tracing in the context of a case study concerning the dynamics of wolf and deer populations on the Kaibab Plateau in the 1920s, a case that has been used as a canonical example of a trophic cascade by ecologists but which has also been subject to controversy. I argue that ecologists have profitably deployed causal process tracing together with other types of experiment to help settle questions of causality in this case. It remains to be seen how widespread the use of causal process tracing outside of the social sciences is (or could be), but there are some potentially promising applications, particularly with respect to questions about specific causal sequences.

1. Introduction

In a classic chapter of a multi-authored edited volume, ecologist Jared Diamond (1986) distinguishes between three types of experiments: 1) lab experiments, 2) field experiments, and 3) natural experiments. Natural experiments are of particular interest for answering questions for which lab experiments and field experiments are not possible or practical; for example, many questions in social science or ecology/evolution are of this sort. But natural experiments typically lack some of the features that provide confidence in causal inferences, such as the ability to regulate variables other than the putative cause. Thad Dunning (2008) suggests that *causal process tracing*, a (typically) qualitative method commonly used in the social sciences, is one way of strengthening causal inferences from natural experiments.

Dunning's intriguing suggestion opens up further questions: how does causal process tracing relate to Diamond's typology of different types of experiment – is it a fourth type of experiment or something else? What can it reveal that cannot be revealed by the different experiment types? Is it relevant outside of social science?

To explore answers to these questions, I'll discuss a canonical case in ecology: the dynamics between wolf and deer populations on the Kaibab Plateau in Arizona in the 1920s. For a time, this case was cited in numerous ecology textbooks as an exemplar of the dangers of predator removal and of a *trophic cascade* (Young, 2002). Trophic cascades have been defined as a situation where "the presence of top trophic-level predators significantly affects herbivores (the next lower trophic level), and this interaction alters or influences vegetation (e.g., species composition, age structure, or spatial distribution)" (Ripple & Beschta, 2005). In the case of the Kaibab, it was claimed that the extirpation of wolves caused deer populations to explode, which led to a reduction in

E-mail address: RLMillstein@UCDavis.edu.

https://doi.org/10.1016/j.shpsa.2019.04.001
Received 17 July 2018; Received in revised form 16 March 2019; Accepted 11 April 2019
Available online 12 April 2019

food for the deer, ultimately resulting in their starvation. Much of the evidence for the Kaibab trophic cascade was based on the work of ecologist Aldo Leopold (1943, pp. 351–366) – known to many philosophers for his Land Ethic – but this was challenged by Graeme Caughley (1970). Thirty-six years later, Dan Binkley, Margaret, William, and Brown (2006) tried to debunk Caughley's debunking and to vindicate Leopold's original conclusions with a thorough re-analysis of the evidence. I argue that Binkley and colleagues deployed natural experiments together with causal process tracing in order to make their case. I conclude that there is, in fact, good evidence for a causal connection between the changes in the wolf and deer populations on the Kaibab that occurred in the 1920s, restoring the canonical case to its rightful place in ecology. I further conclude that using causal process tracing in combination with other experiment types can indeed strengthen causal inferences, especially for particular historical episodes but also for general causation as well.

I proceed as follows. I begin by characterizing Diamond's typology of experiments, followed by a characterization of causal process tracing. I then turn to the case of the Kaibab, first providing necessary background information over the controversy between Leopold and Caughley that Binkley and colleagues were responding to and then examining Binkley et al.'s re-analysis of the case of the Kaibab in light of Diamond's typology of experiments and causal process tracing. I then give discussion of some of the philosophical implications of my analysis and conclude.

2. Diamond's typology of experiments

Jared Diamond (1986) describes his typology of three main types of experiment in the context of community ecology. I will follow Diamond in this characterization, although it should be evident that the three types can be used in disciplines outside of ecology.[1] The basic distinction between the three main types is:

- **Laboratory Experiment** - perturbations are produced by the experimenter in the laboratory.
- **Field Experiment** - perturbations are produced by the experimenter in the field.
- **Natural Experiment** - natural perturbations occur in the field; they are not produced by the experimenter. This includes perturbations due to "humans other than ecologists" (1986, 4).

According to Diamond, in practice, laboratory experiments, field experiments, and natural experiments form a continuum (as my analysis of the Kaibab study will also illustrate). With laboratory experiments, the experimenter regulates the abiotic environment (light, temperature, water, etc.) and the biotic environment. The biotic environment will often consist of communities containing two or a very few species. Common examples in community ecology include organisms kept in bottles ("bottle experiments") or plants kept in greenhouses. Some communities are randomly assigned to receive the treatment (the perturbation, or possible cause under test) whereas other communities do not (the control). The control of the environment is the primary feature of a lab experiment and the one that gives confidence in causal inferences that can be made from the experiment.

With field experiments, the experimenter selects outdoor sites so that, to the extent possible, the sites initially have the same values of unregulated variables The sites are generally adjacent to one another, both for the sake of convenience and to help ensure similarity of sites (this aspect of field experiments will be important later). However, there is no further regulation of the environment beyond seeking similar sites. Some communities are randomly assigned to receive the treatment (e.g., removal or introduction of a species) whereas other communities do not (the control).

Diamond characterizes two types of natural experiments: natural trajectory experiments and natural snapshot experiments. With natural trajectory experiments, comparisons of the *same* community at various times before, during, and after a witnessed field perturbation are performed. The environment is not regulated. Before the perturbation, the site serves as the control; after the perturbation, it can be considered to have received the "treatment". So, the "sites" (really, one and the same site through time) are matched except to the extent they might otherwise change over time. With natural snapshot experiments, comparisons of *different* communities assumed to have reached a quasi-steady state with respect to the perturbing variable are performed. Again, the environment is not regulated. Sites are matched to the extent that it is possible; however, sites are often spatially distant. Sites having experienced a certain perturbation are considered to have received the treatment; sites that have not are considered the control.

With both types of natural experiment, it seems clear that Diamond is emphasizing an analogy to laboratory experiments and field experiments. By showing how they can likewise be characterized in terms of control groups and treatment groups, he shows how causal inferences can be drawn from natural experiments.

According to Diamond, the three types of experiment differ in their merits; there are *tradeoffs*.[2] Laboratory experiments have the highest degree of regulation of independent variables and site matching, both of which give confidence to one's causal inferences, and natural experiments have the least (and so, yield the weakest causal inference), with field experiments falling in between. For practical reasons, natural experiments tend to have the greatest spatial and temporal scale and laboratory experiments the least, with field experiments again falling in between. For both practical and ethical reasons, natural experiments tend to have the greatest range of species and perturbations that can be studied and laboratory experiments the least, with field experiments once again falling in between. With respect to realism,[3] natural experiments are superior to field experiments which are superior to laboratory experiments.[4] Finally, with respect to the ability to follow a causal trajectory, one can do so with laboratory experiments, field experiments, and natural trajectory experiments, but not with natural snapshot experiments.

[1] The sometimes-controversial Jared Diamond might seem like an odd starting point for my analysis, but my reasons for using his typology are fourfold. First, it should be noted that the chapter in which Diamond's analysis appears has been unusually influential for this sort of (frankly, philosophical) piece; as of 22 November 2018, Google Scholar showed 599 citations to the chapter. Second, Diamond is particularly focused on community ecology and so it is appropriate for questions concerning the Kaibab, which are likewise part of community ecology. Third, as I argue below, I think the tradeoffs that Diamond identifies are exemplified in the analysis of the Kaibab. Fourth, the connection between Diamond's typology and causal process tracing is made in a book coauthored by Diamond (with James Robinson), *Natural Experiments in History*. Of course, none of this is to say that Diamond has the final or most complete word on various types of experiments and their tradeoffs, and many philosophers have contributed to a rich literature on these topics. See, e.g., Brandon (1994), Woodward (2003), Odenbaugh (2006), Morgan (2013), Currie & Levi, (forthcoming). A more extensive analysis than this discussion seeks to provide might profitably engage connections between causal process tracing and this literature. Thanks to an anonymous reviewer for pushing me on this point.

[2] See Inkpen (manuscript) for a defense of this claim.

[3] By "realism", Diamond means that the results of the experiment apply to or can readily be extrapolated to any natural community and natural perturbation ("even a single one") (1986, 5). All natural experiments, Diamond implies, are thus completely realistic by definition, whereas lab experiments and field experiments are not, although they can be more or less realistic as a matter of degree.

[4] Relatedly, see Cartwright (2007) and Cartwright and Munro (2010) on the limitations of randomized controlled trials.

Now let's turn to causal process tracing to see how it fits in with Diamond's typology.[5]

3. Causal process tracing

Causal process tracing (CPT) is a bit difficult to pin down; however, David Collier's definition provides a good starting point:

> Process tracing … is an analytic tool for drawing descriptive and causal inferences *from diagnostic pieces of evidence*— often understood as part of a temporal sequence of events or phenomena (Collier, 2011, 824; emphasis added).

> As a tool of causal inference, process tracing focuses on the unfolding of events or situations *over time*. Yet grasping this unfolding is impossible if one cannot adequately describe an event or situation at *one point in time* … To characterize a process, we must be able to characterize key steps in the process, which in turn permits good analysis of change and sequence (Collier, 2011, 824; emphasis in original).

As the highlighted portions indicate, identifying a series of key events or phenomena through time – resurrecting a timeline – is a fundamental part of CPT. It is also typical that the diagnostic pieces of evidence used are quite diverse, including both quantitative and qualitative data. Interviews, written records, etc., are all legitimate forms of data that can be used in CPT. Again, the CPT approach has its origins in the social sciences.

Henry Brady's (2010) discussion of a CPT of the 2000 U.S. presidential election (George W. Bush vs. Al Gore) provides an illustrative example. After the election, the question arose as to whether thousands of votes – perhaps as many as 10,000 or more – were lost for Bush because networks declared Gore the winner of Florida before the polls had closed in western Florida, which is in a different timezone than eastern Florida. To determine the answer to that question, researchers consulted the TV networks to find out when they made their announcements relative to when polls closed in western Florida (10 minutes before). They considered two possibilities, one that the rate of voting in the last 10 minutes was the same as it was throughout the day and the same as it was in the last hour, using data from previous elections as an estimate. Interviews with Florida election officials and review of media reports indicated that there is typically no last minute rush to the polls in western Florida. So, assuming those rates are more or less correct, about 4200 voters could have been affected by the early calling of the election. But what percentage of those voters heard the announcement? Research on media exposure suggests that 20% (840 people) would be very high. But, of course, not all potential voters would be Bush voters; about 2/3 of western Florida voted for Bush (so, about 560 votes). However, a review of past work on the impact of early calls show that they do not deter all voters from voting, e.g., because there are other races at stake. Thus, CPT concluded that the claim for thousands of missing votes lost was not supported.

This case illustrates CPT in its use of diverse sources of data, including written records and interviews, to piece together a timeline: 10 minutes before polls closed in western Florida, media outlets called the race for Gore, an announcement that was probably heard by at most 20% of voters, 1/3 of whom were probably not going to vote for Bush anyway and some of who would still have gone to the polls regardless. Thus, at most 560 voters stayed home rather than voting, and probably less. I note also that in this CPT, past data is used to infer tendencies, such as voter habits, and these are key parts of the analysis.

What is harder to glean from this case is a characterization of a specific *CPT method*. The use of diverse types of data is part of the problem; it's not as though we can specify the statistical methods or inference procedure, as can be done with the various types of experiment. Even the "setup" cannot be given a general characterization the way that one can do with each of the experiment types, which were characterized in terms of treatment groups and control groups with varying amounts of regulation and other aspects. And how the "piecing together" of the different pieces of data occurs is unspecified – and probably has to remain so. It is "detective work". Indeed, examples of CPT in the literature are very disparate: from the analysis of election results just described, to Semmelweis's well-studied discovery of the cause of puerperal fever, to John Snow's study of the causes of a cholera epidemic, to Sherlock Holmes's detective work.

Having said this, researchers have endeavored to characterize CPT methods (see, e.g., Beach & Pedersen, 2013). It would require considerable discussion to present and evaluate this work; my contention here is simply that, by its very nature, much will have to remain unspecified and case-dependent, as even those who seek to describe its methodology seem to admit:

> What empirical material can be evidence of is often case-specific in process-tracing research. The basic point is that the workings of mechanisms often leave different empirical observables in different cases, despite being the same theorized mechanism. Given the very case-specific nature of the evidence implied by the mechanism in different cases, what empirical material counts as evidence in one case is not necessarily what counts as evidence in another. To develop empirical fingerprints that are sensitive to the particulars of individual cases, however, requires considerable case-specific knowledge and expertise (Beach, 2017).

In saying that there is no one univocal CPT method, my intention is not to be critical, but rather, simply descriptive. There is good reason to think that CPT has been extremely fruitful in spite of (or perhaps because of) being somewhat open-ended; the case developed in this paper is meant to be one such example, but there are many others in the literature. In any case, not much hangs on my contention that there is no one CPT method.

I suggest instead that rather than thinking of CPT as a method alongside the experimental methods, perhaps the general class that includes laboratory experiments, field experiments, natural experiments, and CPTs can be considered *cause-finding practices.* If my suggestion here is correct, some questions still remain: What can these different ways of finding causes tell us? Are CPTs useful (and used) outside of the social sciences? Let us turn to the case of the Kaibab for partial answers.

4. The canonical case of the Kaibab

In part inspired by his study of the Kaibab Plateau in Arizona in the 1920s, Aldo Leopold wrote:

> I have lived to see state after state extirpate its wolves. I have watched the face of many a newly wolfless mountain, and seen the south-facing slopes wrinkle with a maze of new deer trails. I have seen every edible bush and seedling browsed, first to anaemic desuetude, and then to death. I have seen every edible tree defoliated to the height of a saddlehorn. Such a mountain looks as if someone had given God a new pruning shears, and forbidden Him all other exercise. In the end the starved bones of the hoped-for deer herd, dead of its own too-much, bleach with the bones of the dead sage, or molder under the high-lined junipers (Leopold, 1949, 130–2).

That quote is taken from his classic essay "Thinking Like a Mountain" in *A Sand County Almanac*, aimed at a general audience. Less poetically, and for a scientific audience, Leopold wrote:

> We have found no record of a deer irruption in North America antedating the removal of deer predators. Those parts of the

[5] See Crasnow (2012, 2017) and Morgan (2012) for other philosophical examinations of causal process tracing.

> continent which still retain the native predators have reported no irruptions. This circumstantial evidence supports the surmise that removal of predators predisposes a deer herd to irruptive behavior (Leopold, 1943, 360).[6]

As mentioned above, the wolves and deer of the Kaibab would come to be seen as the exemplar of the dangers of predator removal and the exemplar of a *trophic cascade*, and the example was included in many ecology textbooks (Young, 2002). Again, the rough idea of a trophic cascade is that a loss of predators leads to an increase in herbivores which leads to a change in vegetation (which can subsequently affect herbivores). Leopold described a formerly stable Kaibab deer herd of around 4000 that began to increase around 1910, with the range showing overbrowsing. According to Leopold, by 1924 the deer herd had increased to 100,000, followed by a famine that reduced the herd 60 percent over two winters. By 1939, Leopold estimated that the herd was down to about 10,000, with a lowered carrying capacity. Leopold cited the loss of predators (cougars and wolves), along with fire control, as events that pave the way for deer irruptions, comparing the Kaibab to similar locations across the U.S.

But the canonical case of the Kaibab was debunked by Graeme Caughley. Caughley's discussion of Leopold is remarkably brief; it is just one section of a much longer paper. Caughley begins by challenging the data that Leopold based his conclusions on, arguing that the records are inconsistent and that some of Leopold's assumptions are arbitrary (the details of his objections need not concern us here, although it's worth noting, as I mention below, that Binkley and colleagues found evidence to support Leopold's data). Concerning these data, Caughley concludes:

> Little can be gleaned from the original records beyond the suggestion that the population began a decline sometime in the period 1924–1930, and that this decline was probably preceded by a period of increase. Any further conclusion is speculative (Caughley, 1970, 56).

After calling into question Leopold's data, Caughley challenges Leopold's *explanation* of the only point of agreement about the data, that there was an increase and subsequent decline in the number of deer on the Kaibab:

> The cause of the eruption is more doubtful than the literature suggests. Increase in deer numbers was certainly concomitant with reduction of pumas and coyotes but it also coincided with a reduction of sheep and cattle. A reported total of 200,000 sheep grazing on the plateau in 1889 had by 1908 decreased to a total of 5,000. Lauckhart and Howard considered that the increase of deer was a consequence of habitat being altered by fire and grazing, and that the reduction of predators was of minor influence (Caughley, 1970, 56; citations removed).

In short, Caughley acknowledges that an increase in deer numbers did coincide with the removal of predators, but states that it also coincided with a reduction in sheep and cattle (who would no longer be competing with the deer for food, allowing deer populations to increase), essentially offering a possible alternative causal explanation for the increase in deer population size. Citing authors who held that the reduction of predators was "of minor influence", fire is added as an additional potential causal factor to grazing. This is the full extent of Caughley's critique of Leopold. Nonetheless, Caughley's brief debunking was seen as sufficiently decisive that the case of the Kaibab was removed from many ecology textbooks (Young, 2002).

Thirty-six years later, Dan Binkley and colleagues sought to debunk the debunker:

> We conclude that Caughley's (1970) hypothesis about the reduction of livestock/deer competition as a driver of the irruption is refuted … The evidence for deer irruptions following periods of reduced predation was consistent for both the 1920s and the 1940s, supporting the idea that predation limits the density of low deer populations, and food limits deer populations (and the absence of aspen recruitment) at high populations (Binkley et al., 2006, 240).

In responding to Caughley (1970), Binkley et al. (2006) focused primarily (but not exclusively) on the herbivore/vegetation part of the trophic cascade, i.e., the claim that an increase in herbivores leads to a change in vegetation. More specifically, they examined the question of whether a loss of wolves and cougars led to an increase in deer which led to a decrease in aspen (deer eat young aspen shoots, preventing aspen regeneration).

5. Binkley et al. (2006) "Was Aldo Leopold Right about the Kaibab Deer Herd?"

The investigation performed by Binkley and colleagues is instructive for the issues that have been raised in this paper because, as I will show in this section, their study can be seen as consisting of three parts:

- a natural trajectory experiment
- two natural snapshot experiments
- causal process tracing (CPT)

To be clear, this is my analysis of their study, not the terms that the scientists use themselves. Nonetheless, I will show that the fit is quite close, allowing us to see the distinctive roles that can be played by natural experiments and CPT in the scientists' overall causal conclusions.

5.1. Part 1: Binkley et al.'s natural trajectory experiment

In one study, Binkley et al. (2006) examined the numbers of quaking aspen of different ages, via a study of the aspens then present across the entire Kaibab Plateau, using tree diameter as a proxy for age. Recall that for a natural trajectory experiment, one performs comparisons of the *same* community at various times before, during, and after a witnessed field perturbation. So, although Binkley and colleagues don't literally observe the trees before and after perturbation, tree diameter allows them to do that by proxy, allowing a study of the aspen population *through time.* Thus, this study can be characterized as a natural trajectory experiment.

Binkley et al. (2006) found that the age structure of aspen on the Kaibab *generally* followed a typical pattern for all-aged forests, with number of trees decreasing exponentially with age. But there were notable exceptions to this pattern: periods when there were *more* aspen than expected (1877–1886 and 1967–1992) and periods when there were *fewer* aspen than expected (1913–1937, *especially between 1923-1927*, the key period in question, and 1953–1962). Importantly, the periods of low aspen numbers correlated with periods of purported deer irruptions – the perturbation – in the 1920s and 1950s.

5.2. Part 2: Binkley et al.'s natural snapshot experiments

In a second pair of experiments, Binkley et al. (2006) examined the ages of quaking aspen then present on the Kaibab comparing aspen within a fenced-in area (no deer) *to* aspen outside that area (deer) as well as comparing aspen within an area protected by dogs (no deer) to aspen outside that area (deer). Recall that with a natural snapshot experiment the researcher compares *different* communities assumed to have reached a quasi-steady state with respect to the perturbing

[6] This quotation and the previous quotation both hint that Leopold saw a number of instances where he thought that wolves and other predators had played a role in the population irruptions of deer and other prey animals. Nevertheless, the focus in this essay will be on the case of the Kaibab, a particular focus of Leopold's as well, although I turn briefly to the question of the more general phenomenon in section 6.

variable. Thus, these studies can be understood as natural snapshot experiments with the deer as the relevant perturbing variable. However, unlike most natural snapshot experiments, the "control" and "treatment" sites were adjacent, meaning that the study was similar to a field experiment in the matching of the sites**.**

Binkley et al. found that "The only successful aspen recruitment during this period was found in areas protected from deer by fences or dogs (2006, 233). Pictures of the fenced in area from 1930, 1942, 1948, and 2003 that Binkley et al. include in their paper show this result very clearly: there are trees within the fenced in area but not outside, except for a couple of young trees in the latest photo. Binkley et al. note that their findings are consistent with an earlier study of forty-one fenced in plots (the one that they inspected was one of these) established in 1927 and examined in the 1930s and 1940s.

5.3. Combining the natural trajectory experiment and the natural snapshot experiments

The natural trajectory experiment and the natural snapshot experiments each provide causally relevant evidence. The natural trajectory experiment reveals that increased numbers of deer and decreased numbers of aspen are correlated *through time.* The natural snapshot experiments, because they involve examination of adjacent sites, give near field experiment-quality evidence that increased numbers of deer are a causal factor for decreased aspen recruitment (again, the only significant aspen recruitment from the 1920s is found in places where there were no deer, adjacent to locations with deer but no aspen).

So, by combining natural snapshot experiments with a natural trajectory experiment, Binkley et al. strengthen their causal inferences; there is evidence that increased deer numbers caused decreased aspen numbers through time. But what about other possible causes? That's where the CPT comes in.

5.4. Part 3: Binkley et al.'s causal process tracing

Binkley et al. (2006) sought out multiple sources of information in order to reconstruct the late 19th-20th century timeline of the Kaibab, looking for correlations between other possible causal factors and aspen recruitment. Their sources included comparative ring-width chronologies for some of the older aspens and pines (to help determine the past effects of climate), interviews with Dennis Lund (formerly with the Kaibab National Forest) and John Goodwin (Arizona Fish and Game Department), published and unpublished records from the Arizona Fish and Game Department, published reports from USDA Forest Services, and secondary sources from peer reviewed journals such as *Journal of Climate.* Recall that CPTs use multiple sources of information, including qualitative sources, to infer the temporal sequence of events or phenomena. Thus, this part of their study fits the characterization of the CPT approach.[7]

As a result of this analysis, Binkley et al. found that increased aspen numbers (which might have led to deer irruption) were not significantly correlated with climate, decrease in livestock (sheep and cattle)[8] but that they were correlated with fire suppression. They found that decreased aspen numbers (which might have resulted from deer irruption) were not significantly correlated with climate, fire, increase in livestock, or logging but were correlated with increased deer numbers. And they found that increased deer numbers were correlated with decreased human hunting of deer, fire suppression, and increased human hunting of predators.

On the latter correlation, essential for defending the causal role for wolves and other predators postulated by Leopold, they found that 5-year periods of low predation by *humans* were all followed by 5-year periods of rapid increases in deer numbers. This supports the general claim that low rates of predation enable the deer population to irrupt. Importantly, other periods with low human predation that lacked intensive predator control did *not* show these irruptions, suggesting that the (non-human) predators were controlling the populations. Finally, they state that "the two major irruptions of the deer population in the 20th century followed periods of major reductions in predation" (Binkley et al., 2006, p. 239).

These findings challenge Caughley's (1970) hypothesis that removal of livestock caused the deer irruption and support Leopold's hypotheses concerning the effects of increased hunting of predators, decreased hunting of deer, and fire suppression.

5.5. Combining natural trajectory experiment, natural snapshot experiments, and CPT

By adding a CPT analysis to the natural trajectory experiment and the natural snapshot experiments, Binkley et al. mitigate the shortcomings of the natural experiments. In natural experiments, independent variables cannot be regulated; Binkley et al.'s CPT rules out other plausible causes of the low aspen numbers in the 1920s and gives some weight to predators over livestock as the cause of the deer irruption. Moreover, CPT rules these out as causes not just at the time in question but through time (late 19th century - 20th century). In sum, by deploying a CPT approach, Binkley et al. leverage the strengths of natural experiments (realism, large spatial/temporal scale) with fewer weaknesses.

However, it should be clear from my discussion in section 5.4 that it does not make sense to call CPT an *experimental* method inasmuch as one cannot identify control or treatment groups. It is not even clear that CPT should be considered a *method* at all (as I argued in section 3). It would be hard to characterize a method that would specify which pieces of evidence the researchers should have looked at. They considered the natural world, various written sources, interviews, etc., but there is no *a priori* guidance that one could give as to which of these sources, or even types of sources, they ought to have consulted. (Moreover, I imagine that their background knowledge played a central role in their choices and that their initial investigations led to checking additional sources that they might not have originally considered). Is "seek out all available sources of causal information in a given timeline that you can think of or learn about" a method? It is a commitment, a practice, but it seems to me too vague to be considered a method. Nonetheless, I believe the analysis of Binkley and colleagues shows it to be a *fruitful* practice, providing information about the presence and absence of various correlations through time that in turn provides (defeasible) evidence for and against various causal factors – evidence that could be obtained in no other way. Thus, as I suggested earlier, we might consider CPT to be a *cause-finding practice.* Since the experimental methods described by Diamond are likewise commitments to seek out available causal information, albeit in more prescribed ways, we can see the experimental methods and CPT as all part of the larger category of "cause-finding practices".

The researchers, unsurprisingly, acknowledge that "uncertainty about causality remains" but also point out that "this level of ambiguity is common in almost all cases involving population ecology, land management, and people" (Binkley et al., 2006, p. 240). For example, they note that "several of the historical reports make passing comments

[7] Similarly, Ripple and Beschta "compiled historical records of wolf kill estimates, by year, from the records of the US Department of Agriculture and obtained information on case studies of ungulate irruptions, by year, for these same western states from Leopold and colleagues" in order to "compare the timing of wolf kills (and ultimate extirpation) with the timing of deer (*Odocoileus spp.*) and elk (*Cervus elaphus*) irruptions to evaluate any temporal patterns in these two variables" (Ripple & Beschta, 2005, p. 614). Thus, their analysis might also be reasonably construed as CPT. It is a measure of the significance of this historical (canonical) case that so much attention has been paid to it.

[8] In these cases, Binkley et al. (2006) cited sources to correct Caughley's (1970) numbers.

about the impacts of rodent grazing" (Binkley et al., 2006, p. 239), which could have affected the vegetation that was available for deer. However, if more detailed information (say, more detailed historical reports) concerning rodent grazing is simply unavailable, then it would be impossible for researchers to determine whether rodent grazing was a significant causal factor over time. My claim, then, is only that the researchers have provided strong evidence for their claims, perhaps as strong as can be expected given the nature of the study and the data that were available to them.

Diamond claims: "Ecologists, like scientists in many other fields, can profit by applying *different methodologies* to *the same system*" (1986, 21; emphasis added). Among other reasons, we achieve a more complete understanding of the system by using different experimental methods because each methodology yields some information that is inaccessible to the others.[9]

The analysis I have given here supports Diamond's claim; moreover, the addition of a CPT analysis yields even more information than just laboratory experiments, field experiments, and natural experiments would have.

6. Discussion

I've argued that Binkley et al. (2006) make use of a CPT analysis in conjunction with natural experiments to settle questions of causality on the Kaibab plateau in the 1920s. CPT has been typically deployed in the social sciences; this raises the question of how common CPTs are in ecology and other natural sciences. Notable here is that Binkley et al. are examining a *particular historical case*.

Sharon Crasnow argues that "specific pieces of evidence produced through process tracing are useful as evidence for singular ["token"] causation" (2012, 665). Binkley et al.'s conclusions fit Crasnow's general characterization; they make two token causal claims: 1) On the Kaibab Plateau in the 1920s, the loss of predators caused a deer irruption and 2) On the Kaibab Plateau in the 1920s, there was a deer irruption that depleted aspen.

This suggests that the CPT approach (perhaps in concert with other types of experiment) may be particularly useful in the historical natural sciences (or natural sciences re-examining a particular historical case) as evidence for token causation, e.g., some parts of evolution, climate science, and paleontology (cf. Cleland, 2001, 2002). Further work might examine whether researchers in these areas are commonly deploying a CPT approach already[10] or whether they might profit from doing so. For example, using fossil data and other indicators of past environments to reconstruct past ecosystems and their trajectories through time (paleoecology) might be profitably understood in terms of CPT, although of course as one's studies recede further into the distant past, certain types of sources (such as interviews and written materials) might no longer be available.

Binkley et al. (2006) also suggest that their case in combination with other cases could be taken as evidence for the general phenomenon of trophic cascades; in other words, they believe that their case contributes to general ("type") causation as well:

> Our results combine with other case studies (for example, Gasaway and others 1992; Krebs and others 2003; Ripple and Beschta 2003, 2004) to indicate that top-down control of food webs is probably not unusual in terrestrial ecosystems (Binkley et al., 2006, 240).

If this is right, then the CPT approach (again, perhaps in conjunction with other types of experiment) could ultimately contribute to type causation conclusions by combining the examination of different cases. To be clear, however, the inference to type causation would not itself be a direct product of experimentation or CPT; rather, the results analyzing particular cases, if deemed to have sufficiently similar causal patterns, could be *generalized* to infer type causation. In other words, the reasoning from the single case to the general case, and thus from token causation to type causation, would simply be inductive.

Relatedly, CPT can be understood as revealing *mechanisms.* Crasnow writes:

> The value of process tracing is thus thought to rest in how it provides evidence for causal mechanisms. Statistical and experimental methods may be able to establish a link between a dependent and independent variable, but they cannot reveal what is in the "black box" – the mechanisms through which the cause brings about the effect (Crasnow, 2017: 7).

Indeed, according to Crasnow, CPT is widely understood this way, with social scientists citing philosophers of science such as Machamer, Darden, and Craver (2000) and Glennan (2002). But I think a bit of caution is in order here. As Skipper and Millstein (2005) argue, the requirements for these accounts of mechanism are more stringent than they might appear on first blush, requiring, for example, that the parts or entities of a mechanism be organized and structured. In the same way that it is doubtful whether natural selection can meet these and other mechanism criteria (Skipper & Millstein, 2005), it is doubtful whether trophic cascades (which can manifest in many different ways with many different types of entities and activities) can be fruitfully understood in terms of these philosophical accounts of mechanisms. The same is true, I suspect, for other causal processes revealed by CPT, such as the case of the 2000 U.S. presidential election discussed above. Nonetheless, the notion that CPT seeks to fill in black boxes between cause and effect (sometimes referred to as "mechanism sketches" in the philosophical literature) seems apt, and there may be cases in which CPT does reveal causal processes that operate in a sufficiently organized and regular way such that it makes sense to see those causal processes as "mechanisms" in the sense espoused by the "new mechanists". (Of course, one might also simply be using the term "mechanism" in a looser sense). In other cases, the causal processes revealed by CPT might be more appropriately described by process accounts such as that of Salmon (1984) or Dupré (2017).

Crasnow (2017) also argues that *narrative* is a core part of CPT, with the construction of a narrative of a particular case contributing to knowledge production. That is, Crasnow argues, the narrative makes causally salient elements of the case coherent. Crasnow's characterization goes beyond the way that many social scientists have portrayed CPT merely in terms of inference and hypothesis testing. In Crasnow's example, "Schultz's account of the Fashoda crisis incident goes on for twenty pages of text … [in which] he tells the story of the crisis in detail, focusing on events, actors, and circumstances of the case" (Crasnow, 2017, p. 11). However, Binkley et al. do not provide this level of detail. Furthermore, although they have reconstructed a timeline (as noted above, a key feature of causal process tracing), the full historical timeline is never actually presented sequentially (first A happens, then B happens, then B happens …). Instead, their primary focus is on answering the question "Was Aldo Leopold Right about the Kaibab Deer Herd?" (the title of the essay) – in other words, was he right about the trophic cascade between wolves, deer, and foliage – to which their answer is "yes". This is primarily a question about whether a particular type of process was operating; this is reflected, for example, in a section that is organized by possible causal factors rather than by the order of events. Thus, what Binkley et al. provide does not seem to fit Crasnow's characterization of a narrative, which is not to deny that they are providing a narrative in some other, perhaps thinner sense. Although the question of narrative is an interesting one, my emphasis here is on the case of the Kaibab as an instance of causal process tracing due to the variety of types of data used and the piecing together of a historical timeline on the basis of that data.

[9] See also Currie (2015) on "methodological omnivory".

[10] See, for example, Wylie (2011) and Forber and Griffith (2011).

7. Conclusion

My aims in this paper have been both specific and general. My specific aim was to illuminate a canonical case in community ecology and the controversy that has surrounded it. To that end, I argued that Binkley et al. (2006) use CPT in conjunction with a natural trajectory experiment and two natural snapshot experiments. Viewing their paper in this light makes clear how their analysis gives good support for the claim that Leopold may have been right about trophic cascade in the Kaibab in the 1920s, i.e., showing that there are good (albeit defeasible) reasons to think that a loss of predators (together with fire suppression) led to a deer irruption which decreased aspen recruitment. This resurrects a canonical case of trophic cascade for ecology as well as a central example for Leopold's claims about the interdependence of species, claims that play an important role in environmental ethics (Millstein, 2018).

My general aim was to show the potential for CPT to add to our cause-finding practices outside of the social sciences. Ecology seems like an unlikely place to find CPT, and yet in this case it was able to enhance the experimenters' natural experiments by ruling out other potential causal factors that (given the structure of a natural experiment) could not be controlled for. No form of experiment could have uncovered these causal factors, which occurred in the past and in some cases are present only in written records or human memory. CPT resurrects the historical timeline, and in so doing provides evidence for the sequence of events present in trophic cascades. Philosophers should be attentive to the possibility for its use in the natural sciences and aware of the benefits it can bring, despite its drawbacks (its open-endedness and its potential absence of controls).

Acknowledgements

Thanks to Chris Young for helpful conversation and for his detailed analysis of the Kaibab Plateau in his book, *In the Absence of Predators: Conservation and Controversy on the Kaibab Plateau*; to two iterations of the Griesemer/Millstein Lab for reading and discussing ideas surrounding this paper; and to audiences at ISHPSSB, PSA, AAAS Pacific, and the University of Calgary for constructive discussion. Thanks also to Andrew Inkpen and the anonymous reviewers for useful suggestions on earlier drafts.

References

Beach, D. (2017). Process-tracing methods in social science. In W. R. Thompson (Ed.). *Oxford research encyclopedias, politics* (pp. 176). Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.
Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods: Foundations and guidelines.* University of Michigan Press.
Binkley, D., Margaret, M. M., William, H. R., & Brown, P. M. (2006). Was Aldo Leopold right about the Kaibab deer herd? *Ecosystems, 9*, 227–241.
Brady, H. E. (2010). Data-set observations versus causal-process observations: The 2000 Us presidential election. In H. E. Brady, & D. Collier (Eds.). *Rethinking social inquiry: Diverse tools, shared standards* (pp. 237–242). Lanham: Rowman & Littlefield Publishers, Inc.
Brandon, R. (1994). Theory and experiment in evolutionary biology. *Synthese, 99*, 59–73.
Cartwright, N. (2007). Are rcts the gold standard? *BioSocieties, 2*(1), 11–20.
Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice, 16*(2), 260–266.
Caughley, G. (1970). Eruption of ungulate populations, with emphasis on himalayan thar in New Zealand. *Ecology, 51*(1), 53–72.
Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. *Geology, 29*(11), 987–990.
Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science, 69*(3), 447–451.
Collier, D. (2011). Understanding process tracing. *Political Science and Politics, 44*(4), 823–830.
Crasnow, S. (2012). The role of case study research in political science: Evidence for causal claims. *Philosophy of Science, 79*(5), 655–666.
Crasnow, S. (2017). "Process tracing in political science: What's the story? *Studies In History and Philosophy of Science Part A, 62*, 6–13.
Currie, A. (2015). Marsupial lions and methodological omnivory: Function, success and reconstruction in paleobiology. *Biology and Philosophy, 30*(2), 187–209.
Currie, A., and A. Levi. (in press). "Why experiments matter." Inquiry: An Interdisciplinary Journal of Philosophy.
Diamond, J. (1986). Overview: Laboratory experiments, field experiments, and natural experiments. In J. Diamond, & T. J. Case (Eds.). *Community ecology* (pp. 3–22). New York: Harper & Row.
Dunning, T. (2008). Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly, 61*(2), 282–293.
Dupré, J. (2017). The metaphysics of evolution. *Interface Focus, 7*(5), https://doi.org/10.1098/rsfs.2016.0148.
Forber, P., & Griffith, E. (2011). Historical reconstruction: Gaining epistemic access to the deep past. *Philosophy and Theory in Biology, 3*, 1–19 e203.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353.
Inkpen, S.A.. Manuscript. "When less manipulation is more? General reflections on experimental trade-offs in the methodology of biology literature.".
Leopold, A. (1943). *Deer irruptions. Natural Resources Committee, Wisconsin Academy of Sciences, Arts, and Letters*.
Leopold, A. (1949). *A Sand county almanac and sketches here and there.* New York: Oxford University Press.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Millstein, R. L. (2018). "Understanding Leopold's concept of 'interdependence' for environmental ethics and conservation biology. *Philosophy of Science, 85*, 1127–1139. https://doi.org/10.1086/699721.
Morgan, M. S. (2012). Case studies: One observation or many? Justification or discovery? *Philosophy of Science, 79*(5), 667–677.
Morgan, M. S. (2013). Nature's experiments and natural experiments in the social sciences. *Philosophy of the Social Sciences, 43*(3), 341–357.
Odenbaugh, J. (2006). Message in the bottle: The constraints of experimentation on model building. *Philosophy of Science, 73*(5), 720–729.
Ripple, W. J., & Beschta, R. L. (2005). Linking wolves and plants: Aldo Leopold on trophic cascades. *AIBS Bulletin, 55*(7), 613–621.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.
Skipper, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 327–347.
Woodward, J. (2003). Experimentation, causal inference, and instrumental realism. In H. Radder (Vol. Ed.), *The philosophy of scientific experimentation. Vols. 87–118*. Pittsburgh: University of Pittsburgh Press.
Wylie, A. (2011). Critical distance: Stabilizing evidential claims in archaeology. In W. Twining, P. Dawid, & D. Vasilak (Eds.). *Evidence, Inference and enquiry* (pp. 371–394). Oxford: Oxford University Press.
Young, C. C. (2002). *In the absence of predators: Conservation and controversy on the Kaibab plateau.* Lincoln: University of Nebraska Press.