A Critical Humanist Perspective on Causal Inference

Nathan Kim

Most interesting academic questions are questions about cause. Academics in the general case don't just seek to understand *what* is happening in some observed event, but *how* and *why* it occurs. These two modifiers force us to examine the causal relationship between observed phenomena, often even when the word "cause" or other terms like "affect" are omitted. In clarifying that causal relationship or describing how one observation affects another, scholars can expose an underlying phenomenon connecting two exterior ones, forming a theory from symptoms.[1]

This essay is an attempt to critique causality itself, a concept scholars usually take for granted in favor of making or critiquing claims of causality. I begin with a discussion of the field of statistical causal inference, which I view as having led the scholarly discussion in this topic for the past thirty years. I generalize its major limitations and issues to argue that any statement of causality is built upon an alternate possibility or potential outcome, and that an infinite portion of these possibilities are foreclosed by the trajectory of history and racial capitalism. In doing so, I attempt to outline the shape of scholary knowledge, the force of cause that motivates much of it, and the dimension of power that modifies how we know cause. Finally, I give two examples to illustrate the application of this line of thought, and conclude with a call to bring this critique into the academy and the world.

What is causality? Most scholars accept what is known as the "counterfactual" model for causality, which defines that A causes B if and only if A were not to happen, then B would not happen. That latter hypothetical scenario is referred to as a *potential outcome*, and forms the crux of a statistical model developed by Jerzy Neyman (1923) and more famously by Donald Rubin (1974 and onwards).[2] In Rubin's words:

> Intuitively, the causal effect of one treatment, E, over another, C, for a particular unit and an interval of time from t_1 to t_2 is the difference between what would have happened at time t_2 if the unit had been exposed to E initiated at t_1 and what would have happened at

[1] Some obligatory caveats here: lots of academics are satisfied with documenting the "what." You can do brilliant, advanced, and theoretically challenging work simply by measuring or describing something – how many of group X is low-income, where is COVID-19 spreading most rapidly, which group won which battle – without attempting to describe some underlying phenomenon that drives it. Cause itself is an infectious question that can find its way into these kinds of topics, but it is worthwhile to simply create an archive or reference material as well.

[2] A general discussion of the history of causality, including causality arguments before the counterfactual model, can be found in Jas Sekhon and J. Daniel Hidalgo "Causality," in *International Encyclopedia of Political Science* (SAGE Publications, 2011).

> t_2 if the unit had been exposed to C initiated at t_1: "If an hour ago I had taken two aspirins instead of just a glass of water, my headache would now be gone," or "Because an hour ago I took two aspirins instead of just a glass of water, my headache is now gone."[3]

Or the corresponding basic mathematical structure:

> Let y(E) be the value of Y measured at t_2 on the unit, given that the unit received the experimental Treatment E initiated at t_1; Let y(C) be the value of Y measured at t_2 on the unit given that the unit received the control Treatment C initiated at t_i; Then y(E) — y(C) is the causal effect of the E versus C treatment on Y for that trial, that is, for that particular unit and the times t_1, t_2.[4]

I'd like to go one step further than Neyman and Rubin. Instead of only arguing that we can prove cause when we prove a difference in potential outcomes, I posit that we only know cause at all, even outside of statistical proofs, when we believe in that difference in potential outcomes. Whether we consciously think about the alternate hypothetical outcome or not, we only ever know a causal effect of event A on outcome B because we believe something else would have happened if event A did not happen. In fact, I'd say these are logically the same sentence.

But there is a glaring issue in the counterfactual approach, which Rubin calls the "fundamental problem of causal inference:" we only ever live in one version of events, one timeline, one strand of history. We never actually know what would happen if event A did not happen, even though we need to in order to show that it causes effect B.

Statistics' main answer to this unavoidable dilemma has been the development of techniques in both experimental design and in observational studies to replicate, randomize, and isolate causal effects. One can only know the single version of events where a participant takes a treatment, so one cannot compare it to the sequence of events where a participant does not receive a treatment. But if a scientist distributes a treatment randomly across many particpants, they don't need to know the alternate timelines for each participant, they can simply take the *average difference in outcomes* across the treatment and control group, and the power of randomness and replication will ensure that any observed outcome is a true measure of the treatment's effect.[5]

Of course, you can't do so in nearly all settings outside of the sciences. There are financial, logistical, and ethical barriers to replicating, randomizing, and controlling treatments, such that the "gold standard" of a randomized experiment is simply not possible. Scholars have tried to compensate for this in observational studies, through finding settings where a condition was near-randomly applied without any researcher's intention, attempting to compare similar participants based on observable characteristics, or using proxy

[3]Donald B. Rubin "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66, no. 5 (October 1974): 688–701, https://doi.org/10.1037/h0037350., p. 2

[4]Rubin "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.", p. 2

[5]It's not really relevant to this paper, but this measurement is called the "average treatment effect," and estimating it or its significance correctly is a large portion of the field of statistical causal inference.

variables better suited for psuedo-experimental analysis as the "treatment" instead of the main variable of interest.[6] But the reality is that these fall short of most questions related to cause, because they require assumptions that are difficult to justify and often cannot be met. Even when they are successfully applied, they are regretfully but understandably looked down upon in their argumentative power compared to experimental designs, because there may be thousands of unobserved confounders; after all of your efforts, you may still be studying a case of correlation, not causation. And of course, the field of causal inference as it stands is almost completely irrelevant for the humanities and many social sciences simply because the kind of data dealt with does not allow a single statistical test to be performed. This is not to say that causal assertions are not made in these fields, because as I began this essay with observing, statements about cause form the core of most academic work, but supposed "foolproof" methods for asserting causality are simply not applicable to much of the academy.

With this inescapable truth, I want to return to my above observation that cause is invariably known through potential outcomes, even without any researcher's intent. How can that be possible? Despite having no way to know what happens in that alternate timeline, and no way to compute the average of many different randomized conditions and outcomes, how can we infer cause? The short answer is that the counterfactual is imagined. "Imagined" might seem like a strange word, because we mostly don't consciously think of that alternate reality when we assert a causal relationship. I choose it because the way we know cause in the absence of a strict scientific proof is almost by definition subjective, relying on all of our qualitative and malleable leanings in its construction. Our assertions for cause depend on our belief that a different effect would be produced if a situation began with a different cause, and like all of our beliefs it has been formed by a host of personal experience and social forces.

To clarify, the kind of imagination involved in the causal assertion is not a momentary impulse we have. As I said above, we don't think of that alternative reality when we make a causal assertion, and it doesn't arise in our subconscious in that instant either. Instead, over time a belief in what could have been becomes ingrained into our subconscious and into the fabric of our current reality, through how we know history and time, through the images we form of places around us, through the demographics and built realities of our current social realities. These forces serve to memorialize, preserve, and excavate certain forms of knowledge, and other times to create gaps and let the ebb of time wash history away – all in line with the hierarchy of power in our current social formation. Thus, just as how the field of statistics often refers to the fundamental problem of causal inference as a "missing data problem," the erasure of these possibilities might be seen from the perspective of the humanities as an archival problem of what kinds of knowledge are preserved for future scholars to study. But more than being "just" a methodological problem, the fundamental problem of causal inference is just as much a problem at large of the kinds of people, possibilities, and futures are erased through the forces of war, genocide, oppression, racism, and the dimension of power in general. The alternate possibilties available to us to make the causal assertion are built over time through an uncountable number of agents that form structural forces, and an equally uncoutable number of these possibilities are lost forever as the course of history inevitably and invariably forgets what could have been.

In her landmark work *The Intimacies of Four Continents*, Lisa Lowe refers to the problem of "what could have been" as the *past conditional temporality*, or a "space of a different kind of thinking, a space of productive attention to the scene of loss, a thinking with twofold attention that seeks to encompass at once the positive objects and methods of history and social science, and also the matters absent, entangled, and

[6] More commonly referred to as natural experiments, propensity score matching and weighting, and instrumental variables.

unavailable by its methods."[7] Drawing from Stephanie Smallwood, David Eng, and David Kanzanijian, Lowe views violent colonial removals from the archive alongside the archive's disciplinary and geographic segmentations, all of which she attempts to think across in her monograph. Saidiya Hartman famously brings an analogous concept forth in *critical fabulation*, "a critical reading of the archive that mimes the figurative dimensions of history ... to tell an impossible story and to amplify the impossibility of its telling."[8] We may think of the many scholars and writers in science fiction and speculative fiction, historians recovering lost or missing archives, and those in ethnic studies and cultural studies participating in that project of critical fabulation. They clarify those alternative possibilities that our current hegemonic racial formation tries hard to make impossible to imagine. In doing so, they not only take on the project of reckoning with the scale and extent of loss but also the impossible task of describing how loss has affected the present.

This essay does not attempt to join that project through archival recovery or speculation, but instead through applying these scholars' line of thought to the domain of causal inference and potential outcomes. I want to focus on what I will call the *specter of alternate possibility* – that any causal assertion we make today has been and is being shaped by realities that are no longer possible, which at some point in history served a reference for a causal assertion to be made, but only haunts our causal assertions today. *Specter* is partially inspired by Grace Cho's *Haunting of the Korean Diaspora*, a history of the Korean War and how its aftermath is felt through both material reality and ghostly silences passed down for generations. I also use it to bring forth Marx and Engel's famous assertion from the *Communist Manifesto* that a "specter is haunting Europe – the specter of communism," as well as Derrida's revival of the metaphor and the coinage of "hauntology" in *Specters of Marx*. What I want to gesture towards with these references is not only the permanence and power of loss, but alsothe paradoxical *presence* of loss in how it continually shapes our reality today, and to locate the type of loss most relevant to the specter in the flow of power within empire and racial capitalism.

To bring us back to a less lofty point and to argue for the ghostly power of the specter, I want to consider some examples of the specter's haunting. Alice Amsden writes in *Asia's Next Giant* (1992) that Korea fell to Japanese colonization because it was "weak," relying on and reifying a Western assessment of precolonial Korea as such in comparison to South Korea's later period of industrialization. But how do we know – what would have happened if Korea was "strong?" And might the problem instead be located in Japan's imperial and capitalist ambitions, or in a lack of allyship from troubled neighbors Russia and China? The conclusion that Korea was conquered because of some weakness is acceptable to Amsden because Japan having an alternate set of goals is not a valid alternate possibility. One cannot make a meaningful statement that Korea was conquered because of Japan's imperial ambitions because the ambitions are a plainly obvious fact, an inexchangeable prior over which other alternate possibilites are considered. We never really know the outcome of the alternate possibilities listed above, for we only have the timeline in which Korea was "weak," but paying heed to the specter of alternate possibility and asking *what if* forces us to reexamine the assumptions over which these kinds of causal conclusions are made, and how the causal conclusion reinforces the assumption that prefaced it.

Finally, though I began the discussion of the specter by considering the cases in which randomized experiments are not possible, I want to conclude by arguing that the social construction of cause absolutely

[7] Lisa Lowe *The Intimacies of Four Continents* (Duke University Press, 2015).

[8] Saidiya Hartman "Venus in Two Acts," *Small Axe: A Caribbean Journal of Criticism* 12, no. 2 (June 2008): 1–14, https://doi.org/10.1215/-12-2-1.

haunts experimental settings. To give a small example, we might conduct a "gold standard" of the randomized experiment to find that individuals spending more time on their cell phones as children exhibit diminished social skills, and conclude that phones cause harm for child development. The experiment may be entirely valid in its conclusions and methodology, and it may provide a well-qualified argument for many parents to lower screen time for their children. But we limit ourselves in accepting its conclusions forthright as a statement of phones as an abstract concept affecting health. What if the past forty years had yielded a different kind of technology, one that did not alienate us through endless streams of content and advertising? Accepting this experiment's conclusions of phones in the abstract or universal case causing harm for child development is only possible if we forget that alternative possibility, and conversely we foreclose a revival of that alternative present in accepting this experiment's conclusions as a reflection of the natural world.

These are brief and trivial examples, but I hope they illustrate the infectious nature of the specter, in both settings where we make "unprovable" assertions of cause and when we abide by the "gold standard" of the randomized experiment. The forms of the arguments involved in these examples are ubiquitous in scholarly knowledge, hinting at opportunities to rethink causality everywhere. And just as how the specter of alternative possibility is born through losses both in the archive and in the world at large, to release the specter from its quest of haunting requires us to reckon with both scholarly knowledge and to move towards a new world built on care, preservation, and remembrance. Making claims of causality requires us to imagine a world we don't know based only on the world we do know, which is forged by the unending march of racial capitalism; perhaps we can move towards a better world by imagining otherwise.