Holistic Causal Learning with Causal Graphs: A Credible Method for Study Design and Preregistration in the Social Sciences

Robert Kubinec[1,*]

October 30, 2024

Abstract

While research designs in the social sciences have employed increasingly sophisticated methods to control false positive rates, there is still substantial debate about the merit of pre-registration and other recent open science reforms. In this paper, I present a method for preregistering causal graphs and employing the metric of entropy, and in particular Jaynes' theory of maximum entropy, to propose a holistic way of measuring the contribution of a research study. To demonstrate the method's utility, I show how recent research in both COVID-19 and political authoritarianism can be fruitfully understood using causal graphs and entropy. Additionally, I provide R code to enable researchers to compute these metrics, helping them prepare for various outcomes and learning approaches for the purpose of preregistration.[1]

[1] University of South Carolina

[*] Correspondence: Robert Kubinec <rkubinec@mailbox.sc.edu>

[1]A reproducible version of this paper with code is available at https://github.com/saudiwin/causality/blob/master/prereg_causal_graph.qmd . I thank David Waldner, Christopher Winship, Michael Poznansky, Kevin Munger, Arthur Spirling, Andrew Gelman, Kosuke Imai and participants in Polmeth Asia 2023 for helpful comments on this manuscript.

The past decade has seen remarkable changes in how social scientists are expected to conduct research studies, especially when the study involves experiments or aims for causal inference more generally. In order to combat questionable research practices (QRPs) that undermine statistical findings (Schneider et al. 2023), open science reformers have proposed pre-registration as a form of binding commitment device that (in theory) forces researchers to report all relevant analyses (Scoggins and Robertson 2023). Indeed, some political science journals have recently made pre-registration mandatory,[2] leading to an increasing adoption of the norm that pre-registration is necessary for experimental inference and possibly observational (or quasi-experimental) inference as well.

Despite this widespread adoption, evidence of the effectiveness of pre-registration remains scant and has been questioned recently (Brodeur et al. 2024). In one notable recent episode, Open Science Foundation authors used results from outside their own pre-registered study of the effect of pre-registrations (Bak-Coleman and Devezer 2024), which the journal asked them to retract (Protzko et al. 2024). The basic logic of pre-registration as a "bind the hands" device has been further questioned as it may only shift the selection away from null effects and other undesirable research outcomes to a different stage (Schakenberg, Little, and Montgomergy 2024). For these reasons, while pre-registration is becoming a growing norm, it is also increasingly under fire for failing to live up to its promises to make science a fairer, more transparent, and ultimately more credible enterprise.

In this paper, I propose a new method for understanding preregistration as involving a registration of the state of the world of researcher knowledge that is expressed in the form of a causal graph. This method builds on recent advances in social science research design that is increasing the capability of tools for applied researchers to diagnose issues and construct robust theories (Blair, Coppock, and Humphreys 2023; Humphreys and Jacobs 2023). With a preregistered causal graph along with suitable priors over the relationships between vari-

[2]For more information, see https://tompepinsky.com/2021/01/16/on-requiring-pre-registration/.

ables (Humphreys and Jacobs 2023), authors could then employ metrics based on entropy, in particular Jaynes' (2003) theory of maximum entropy, to determine the level of causal knowledge that a particular study would contribute. Importantly, knowing the amount of prior causal knowledge about a causal system would permit robust post hoc exploratory inference of the kind that is often done in the social sciences.

Based on Jaynes' theory, I provide a formal basis in this paper for two forms of causal learning: the more common type I causal learning (searching for new relationships) and the less common yet equally as important type II causal learning (learning from nulls as in Alrababa'h et al. (2022)). Given an open-ended preregistration that allowed for both types of causal learning, study authors could then determine with much more precision the level of causal knowledge obtained without requiring a specific set of analysis steps in the document.

I propose that we can accomplish this objective by applying Jaynes' (2003) theory of maximum entropy to the evaluation of causal systems—specifically, by calculating the entropy of the joint distribution of variables in a causal graph. Entropy is a framework widely used in statistics to represent the relative amount of information present in a random variable (Shannon 1948). Entropy is a characteristic of a random variable's distribution; the flatter (more uncertain) the distribution, the more entropy exists. When applied to causal graphs (Pearl 2000), the theory of maximum entropy requires us to begin with minimal assumptions and always seek to maximize our uncertainty given these set of minimal assumptions.

The aim of this method is to encourage researchers to make causal graphs central to their research by directly incorporating them into pre-study planning such as preregistrations. I further provide helpful guidelines for researchers to employ causal graphs in realistic settings where there is disagreement over the true causal process. I argue that even in cases where the definition of the causal system is uncertain, framing research designs in terms of causal graphs and employing Jaynes' theory of maximum entropy can provide helpful baselines for reasoning about optimal interventions for causal learning and avoid disagreements about

the "correct" analyses to conduct following an experiment. For this reason, the method of entropy I propose is related to scholarship about obtaining real-world learning out of biased data distributions (Spirling and Stewart 2022; Huang 2024; Dorie et al. 2016; Slough 2023; Berinsky, Druckman, and Yamamoto 2021).

I show how this framework can permit diverse forms of causal learning from case studies of two important cross-disciplinary research topics: the efficacy of COVID-19 vaccines and the relationship between a country's oil wealth and its level of democracy. I show through the case studies how observational, experimental, and mechanistic (qualitative) research designs can each contribute causal knowledge in varying amounts depending on the pre-specified causal structure. Given a fixed research budget, I even show how this method can even determine an optimal research design in terms of maximizing learning about causal systems given minimal assumptions.

The method of entropy developed in this study has applicability to the decision-making calculus involved in research design, the evaluation of the amount of learning from different studies, and the meta-analytic task of summarizing the amount learned from studies in a given field. The method can be estimated using R code provided in the appendix and makes use of new frameworks for power analysis and research design (Humphreys and Jacobs 2023; Blair, Coppock, and Humphreys 2023) to enable social scientists to be able to make valid and robust decisions about what kind of studies to pursue given constraints, as well as to defend their decisions to pursue a certain path of research even if they cannot guarantee the causal identification of relations between single variables in a given study.

A Theory of the Learning of Causal Systems

To make preregistrations optimally useful, they must be understood within a larger theory of inference. Existing frameworks for preregistration tend to follow a Popperian type of logic

in which the scientist derives hypotheses independent of data collection (Popper 1963), and experiments are attempts to refute or confirm these previously derived hypotheses. While it is not my aim to dispute this common understanding of the scientific process–though it has notable critics (Agassi and Institute of Philosophy, Russian Academy of Sciences 2022)– it is rather much easier to point out that in social scientific practice, the Popperian ideal is rarely if ever achieved (Caldwell 1984). Our theories are too imprecise (Frankenhuis, Panchanathan, and Smaldino 2022), our measures are too noisy (Flake and Fried 2019), and our data collection is too difficult to allow for tests that can be seen as fully confirming or disconfirming for most studies we perform (Gelman 2015).

The traditional view of preregistration cited earlier has its aim of bringing science closer to the falsificationist goal. Similarly, the recent criticism of the movement alleges that preregistration norms have failed to stop p-hacking and other QRPs. However, preregistration can serve multiple aims, not all of which have to do with falsification per se. After all, much of social science research is exploratory (Swedberg 2020; Gerring 2012), either because existing theories do not cover the empirical data, or because the theories are too imprecise as to provide clear predictions before observing the data. Given this difficulty, preregistration of causal graphs provides a path forward for permitting exploratory analyses that are robust and credible but also do not require researchers to invent post-hoc theoretical justifications for their decisions.

There is a vast and growing literature on causal graphs, much of it stemming from the Pearl (2000) revival of interest in causal systems. While social-scientific research designs often assume a binary treatment and a binary outcome, social science theories can make use of many variables, and causal graphs are able to incorporate much more of this complexity than the potential outcomes framework can permit. This flexibility is very important for avoiding the previously described issues of preregistrations as p-hacking very often involves subgroup analyses and other exploratory approaches that incorporate additional variables.

This paper is certainly not the first to show that causal graphs can help advance research goals in difficult areas. These type of structural causal models have been increasingly applied to diverse phenomena across the social sciences (Blackwell 2013; Liu, Wang, and Xu 2024; Waldner 2015). My intention here is to incorporate causal graphs in the same manner as Humphreys and Jacobs (2023) recent work in terms of a tool for organizing and understanding theories which are subsequently tested with data. I defer the reader to their treatment for a more thorough understanding of the underpinnings of causal graphs and their relationship to individual research designs.

One of the most important objections to relying on causal graphs is that it is difficult to know when a causal graph is correct, or in the parlance of the literature, complete (Waldner 2015). Theoretically, a complete causal graph would permit a scholar to determine causal relations from purely observational data; practically, if we knew true causal graphs for social processes, we would have little need of further research. A true causal graph is a multidimensional analogue of the frequentist idea of a population parameter: while we want to find this quantity, we can never know for certain what it is except in expectation.

Nevertheless, I argue that a causal graph can help even in situations where substantial uncertainty exists over the true causal graph, as I explain in this paper. So long as the analysis can limit itself to a specific set of variables—which is in many cases a requirement for any kind of research study—then a causal graph, even a very simple one, can be derived. Given the full range of potential causal graphs, even up to the set of all bidirectional graphs, the entropy metric can then be used to determine how much causal learning is possible depending on the type of research design that is pursued.

Entropy is a concept with a long history in statistics, though in this paper it is used primarily as a way of making nuanced statements about the sum total of uncertainty in a causal system. In general, entropy describes the decay of a system, such as gas molecules moving farther and farther apart to fill a sphere. Statistical, or Shannon, entropy applies the same concept to

probability, providing a measure of the "information" in a random variable (Shannon 1948). Shannon entropy H is defined as a simple formula applied to a distribution of N probabilities that cumulatively sum to 1:

$$H = -\sum_{n=1}^{N} p_n \log p_n \tag{1}$$

The formula in (1) is unfortunately not intuitive, in part because the units of entropy are in logarithms. Because I am interested in entropy as a framework rather than with a particular empirical application, I use the more intuitive logarithmic base of 1.01:

$$H = -\sum_{n=1}^{N} p_n \log_{1.01} p_n \tag{2}$$

A base of 1.01 means that every unit increase in entropy equals a one percent increase in entropy. Figure 3 plots entropy calculations for probability distributions with varying levels of uncertainty or spread-out-ness. What is important to note is that all of these distributions have the same expected, or average, value, but are nonetheless very different statements about underlying uncertainty. Roughly speaking, the uniform distribution has 100 percent more entropy than the normal distribution, which has 30 percent more entropy than the student's T and Laplace distributions. These plots show why entropy is a powerful heuristic: it captures our sense of how certain we are of the empirical possibilities underlying a distribution of probability that is independent of the form of the distribution. If we know nothing about a process, we can assume a uniform distribution that leaves probability mass on all possible outcomes, or what is known as the maximum entropy distribution per Jaynes. But if we know more about how a process operates, we can considerably reduce our uncertainty (and hence entropy) by choosing a more specified distribution.

It is worth noting that entropy and variance are both statements about the uncertainty in

a statistical distribution but are not interchangeable. Changes in the variance of variables, such as through Bayesian updating, provide similar information to entropy when considering repeated observations of the same random variable (Humphreys and Jacobs 2023, 327). However, it is important to note that in general entropy is a completely different measure than variance. Entropy does not carry the units of a random variable as the variance does, which permits entropy to be used across variables of different distributions (or even unknown distributions) in the same causal system. Variance is also defined as the second moment of a given distribution, while entropy is not a moment of any kind. Conceptually, entropy combines information from different moments, including variance, skewness and kurtosis, which is the reason it is often the basis for general measures of model fit like the (Wanatabe-)Akaike information criterion (AIC).

Finally and most importantly, theories and lemmas built on entropy, such as Jaynes' maximum entropy considered here, do not hold when variance is the underlying measure. The relationship between variance and entropy varies significantly across distributions (Mukher jee and Ratnaparkhi 1986), which limits the utility of relying on variance as a heuristic for entropy.

While entropy has been applied successfully to many statistical problems, my intention in defining it here is to think of it as a way to understand the relative value of the different research designs, whether they be experimental or observational. The maximum entropy principle provides clarity about how we can maximize knowledge while avoiding overconfidence. Jaynes (2003) defines the maximum entropy principle as always preferring a distribution of higher entropy conditional on including all known facts in the distribution. For example, suppose we wanted to predict stock market prices. Lacking any special knowledge of stock prices, we would want our uncertainty to reflect the fact that all we have to analyze are the movements of individual stocks over time–we would want to maximize entropy or uncertainty, given the data we have. But if we knew that the Federal Reserve intended to

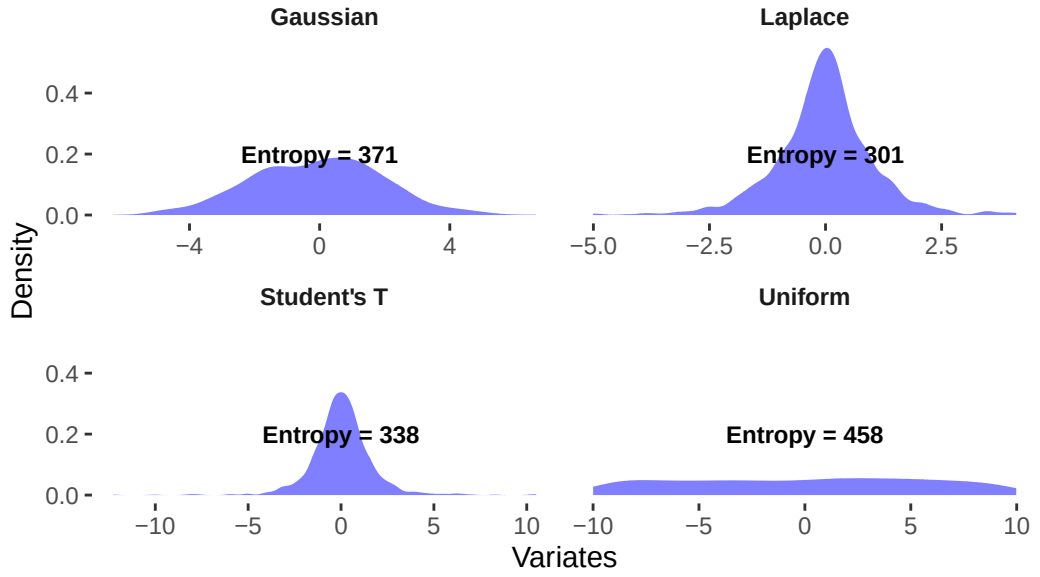

Figure 1: Entropy Calculations Based on Empirical Densities of Statistical Distributions

increase interest rates, we could include that information in our model and consequently obtain a lower entropy distribution.

Ultimately, the goal of the social sciences should be to reduce entropy whenever possible in terms of our understanding of how the social world operates. If we have a more certain knowledge of the distribution of outcomes, we can state with reasonable confidence that our knowledge is increasing (Gerring 2006). To do so, we have to produce new propositions that explain human behavior and allow us to make judgments about what is more or less likely to occur. From a Bayesian point of view, we could re-state this problem as meaning that we should always prefer the research design that increases our knowledge relative to our prior, even if the knowledge we obtain has residual bias (Little and Pepinsky 2021).

In other words, we want to learn new facts about the world such that we reduce our entropy in understanding causal systems. At the same time, we want to maximize entropy given what we know to reduce blind spots and over-confidence. Causal inference involves striking this delicate balance between assuming too much and assuming too little (Clarke and Primo

2007).

Based on the principle of maximum entropy, I consider two ways that changes in entropy can map onto learning about causal systems. Generally, we tend to think of causal learning as involving reducing entropy, that is, reducing our uncertainty and adding structure to our understanding of the world. I call this type I causal learning. At the same time, if we are overconfident in our causal knowledge, we may need to instead inflate our uncertainty, or what I term type II causal learning. Type II causal learning relates to the growing body of literature about the importance of making inferences concerning "null effects" in experiments (Alrababa'h et al. 2022).

To use entropy to understand causal systems, I employ Pearl's directed a-cyclic graphs. My intention is not to suggest that Pearl's theory is the final take on causality, but rather that causal diagrams are flexible ways of encoding a wide array of assumptions about causality. As such, they are a helpful starting block for comparing very different representations of causality. There are existing applications of entropy to causal graphs, but the aim in the literature is to uncover hidden confounders given a set of observed data as opposed to making larger statements about research design (Kocaoglu et al. 2020; Tee, Parisis, and Wakeman 2016; Wieczorek and Roth 2019).

It is important to note, however, that by employing causal graphs I am making the implicit statement that at least some assumptions are necessary for causal inference. This approach contrasts with the so-called "barefoot" experimentalist approach in which experiments can be applied iteratively to learn causal relations without the need for theory or consideration for the larger causal system (Green and Gerber 2003; Deaton and Cartwright 2018; Ashworth, Berry, and Bueno de Mesquita 2021). My employment of causal graphs is not intended to be a statement about this debate as my framework cannot be fruitfully applied without some structure to describe a causal system.

For these reasons, it is important to note that the minimal causal system is a causal graph

of at least two nodes. A bivariate graph with only a single independent variable and single dependent variable is a perfectly acceptable starting place (and all the calculations in this paper apply) although the true power of causal graphs generally comes from understanding more complex systems. However, care must be taken when determining the set of nodes over which the causal system is to be defined. As is well-known, the complexity of a directed graph increases at a highly non-linear rate as each additional node $n \in \{1, 2, ...N\}$ allows for $N + 1(N) - N(N-1) = 2N$ additional edges (connections) between nodes. At the same time, taking into account the principle of maximum entropy, we would only want to exclude edges from a causal graph if such an assumption is uncontroversial. If there is residual uncertainty, we should include an edge between nodes in the graph. Thus the definition of a causal system involves the selection of a set of nodes, and research on the causal graph primarily focuses on adjudicating the types of edges (connections) between said nodes.

Once the analyst has chosen the set of variables which pertain to the research question, it is possible to calculate entropy of any number of possible causal graphs. I now formally define the causal entropy measure as the Shannon entropy $H(\cdot)$ of an ordered set of variables (nodes) $\{x_1, ...x_N\}$ that can be represented by a causal graph v which meets all of Pearl's requirements: it includes a set of directed edges $e \in E$ between each variables x_i in the graph and is acyclic. Because the graph is ordered, there is necessarily at least one terminal node that represents the outcome that the graph seeks to explain. We can then take the Shannon entropy $H(\cdot)$ of the joint distribution of these variables (indexed by $i \in N$) with respect to a given terminal node or outcome, which can be denoted $P(v)$, following Pearl's notation:

$$H(P(v)) = -\sum_{i=1}^{N} P(x_i|pa_i)P(pa_i)\log P(x_i|pa_i)P(pa_i) \tag{3}$$

Where the notation $P(x_i|pa_i)$ indicates that each component of $P(v)$ is the conditional distribution of each variable x_i in V with respect to the set of its ancestors pa_i that are

causally relevant to x_i. Due to the Markov property of causal graphs (Pearl 2000, Ch. 1), we only need to consider the immediate set of ancestors as causally relevant, which significantly simplifies the construction of conditional probabilities for a given x_i. Because $P(v)$ is a joint distribution over all such relations, it meets the requirement that the probabilities of all of the causal relations sum to 1.

Formally, we can then consider research designs as representing different joint distributions, such as $P(v)$ and $P(v')$. We can state that a research design that results in $P(v)$ creates more type I causal knowledge than a research design that produces $P(v')$ iff:

$$H(P(v)) < H(P(v')) \tag{4}$$

Conversely, to understand type II causal learning, we need a measure of cross-entropy, or comparing the entropy of two distributions. We can suppose first that we have a maximum causal entropy distribution $P(q)$ for a given causal graph which represents a state of maximum uncertainty about the causal process. We also suppose that the current state of the research field accepts a reduced causal entropy distribution $P(v)|H(P(v)) < H(P(q))$, but this distribution is based on flawed data or assumptions, resulting in overconfidence.

We could obtain type II causal learning if we were to implement a replication study $P(v')$ that results in higher entropy than $P(v)$, bringing us closer to the maximum entropy of $P(q)$. With the maximum entropy distribution $P(q)$ as our baseline, the amount of type II causal learning we obtain depends on how much the Kullback-Leibler divergence $D_{KL}(P(q) \parallel P(v))$ decreases from the maximum entropy distribution following the replication $P(v')$:

$$H(P(q)) + D_{KL}(P(q) \parallel P(v)) > H(P(q)) + D_{KL}(P(q) \parallel P(v')) \tag{5}$$

In other words, if our replication exercise $P(v')$ raises our uncertainty about the causal graph

$P(v)$ relative to the limiting case of the maximum entropy distribution $P(q)$, then we can say that we have obtained type II causal knowledge in that we are no longer overconfident about the causal process. This notation shows how we could quantify such an increase in knowledge; it does not specify under what standards we would believe $P(v')$ to be more credible than $P(v)$. As mentioned previously, entropy is about understanding the relative level of type I or type II causal learning from a given inference but is not a criterion for deciding which inferences are based on credible information. What we gain from this framework is primarily a way of comparing different causal graphs with the same number of nodes yet varying types of edges between nodes.

Estimating type II learning requires a maximum entropy distribution as a the baseline, but these are relatively easy to derive for most cases. As was shown in Figure 1, the uniform distribution is the maximum entropy distribution if there is no other structure or information to be taken into account. In the case studies that follow, I will use the uniform distribution as the maximum entropy distribution, or what I will refer to as the "null" graph. The Normal or Gaussian distribution is the maximum entropy distribution for any continuous variable with finite variance (Jaynes 2003), and so it can serve as a null graph in the case in which units are not discrete probabilities.

One complication is that entropy is only defined over discrete variables. However, it is straightforward to calculate the entropy of a continuous variable through a binning procedure, and the measure is available via a wide array of statistical software packages (Hausser and Strimmer 2021). In the appendix, I provide a tutorial with R code showing how to calculate these quantities given different causal graphs. I further show in the appendix how the framework of Humphreys and Jacobs (2023) can model causal systems in which there is uncertainty in the probability of edges between nodes via the Dirichlet distribution; this methodology permits statistical inference on changes in type I and type II causal learning for different research designs, though it also requires the specification of prior distributions

over nodes in the causal graph.

Once we have defined these forms of learning about causal systems, we can further define an optimal research design q^* as that which maximizes a research utility function for entropy minimization (type I causal learning) given a prior graph $P(v)$, a set of plausible research designs $Q = \{P(v_1), P(v_2), P(v_3)...P(v_N)\}$, and a mapping of costs for each research design $C = \{c_1, c_2, c_3...c_N\}$ where we select the research design that maximizes type I learning given a budget constraint B:

$$q^* = \arg\max_{q \in Q} U_i(B - c_q - (H(q) - H(P(v)))) \tag{6}$$

$$\text{s.t. } U_i(\cdot) > 0 \tag{7}$$

Preregistration of Entropy and Causal Graphs

To make this method useful, however, it is not necessary to solve for the optimal research design. It is only necessary to derive such a graph–or distribution of possible graphs–prior to undertaking a given study. Once the prior causal graph is known–ideally with probabilities over possible edges–then exploratory inference is permissible as false positive rates are known. If the causal graph is preregistered—even a causal graph with substantial uncertainty—then any post-study analyses of the study's data will be valid so long as they are shown relative to the preexisting information included in the preregistration.

To demonstrate this point, consider Gelman and Loken's analysis of the so-called "garden of forking paths" that motivated much of the movement for preregistration (Gelman and Loken 2013) in which we are concerned about obtaining an accurate test statistic $T(y)$ for a given dataset y. Gelman and Loken's concern was that the quantity reported in many papers is in fact a function of the actual data obtained (i.e., statistics to report depended on

the outcome of the experiment), that is, $T(y, \theta(y))$ where $\theta(y)$ is a function that determines which statistics to report. Gelman and Loken's point is that test statistics are invalid if they are a function of the data that an analyst actually obtains. Preregistration is supposed to remove this endogeneity by stipulating the set of tests chosen θ ahead of time: $T(y, \theta)$ so that it has no relationship to the data an analyst obtains.

Gelman and Loken's concern with this endogeneity in how statistics are reported is that it can artificially reduce uncertainty, or what he calls the "researchers' degrees of freedom problem" (p. 1). Essentially, if the test selection function $\theta(y)$ is optimized to find the most interesting finding–or even if it does not but still depends on the data obtained, then the test statistic is evaluated against the null hypothesis, that is, supposing that nothing is known a priori about $T(y)$. In this case, the reported p-value is based on a faulty understanding of the sample size given that there is information about the outcome y in the choice of test statistic $T(y, \theta(y))$–that is, if the outcome had been different, this particular test statistic would not have been chosen, and thus the sampling distribution of the test statistic is invalid.

However, if we have preregistered the causal graph that we believe can generate y, denoted $P(v_y)$, following the notation above, then we can always back out a more informed baseline than the null hypothesis–we know the inputs to $\theta(y)$ because our selection of test statistics is aimed at minimizing entropy and maximizing a posteriori causal learning: $\theta(y) = \arg\min H(P(v_y))$. This informative baseline—the state of our knowledge about the causal process—permits exploration without reaching misleading conclusions. We no longer have to pretend that we selected our tests $T(y)$ with no attention paid to the data at hand.

For example, suppose a scholar does an experimental study of voting behavior using a vignette experiment. The outcome y is the proportion voting in the election and $T(y, D)$ are the analyses that the scholar could report examining the relationship between the treatment D and the outcome y. Suppose that the scholar recorded some preregistered analyses θ

and reports these in the paper for $T(y|D,\theta)$. These test statistics are evaluated against the null hypothesis and are valid as they are independent of the observed data. However, the analyst found a surprising treatment interaction between gender, denoted G, and voting y, but this analysis was not preregistered: $T(y|D,G,\theta(y))$. Because the analyst did not know the relationship would be surprising until after observing the data, the decision to choose to report it based on the observed dataset could result in misleading inferences if the analyst acted as though this had been their plan all along.

However, if G was in the causal graph—even if the preanalysis plan did not include a specific test for G–then the amount learned from the new analysis can be understood as the difference in entropy of the null (or pregistered) graph $H(P(v_{yt-1}))$ and the entropy of the causal graph following the experiment $H(P(v_{yt+1}))$. This quantity is valid regardless of the number or type of analyses that are reported in the article because it entails making a holistic comparison of causal graphs inclusive of all variables that contribute to y. The null hypothesis for any particular test statistic $T(y)$ could be adjusted based on the preexisting level of uncertainty or entropy so that truly surprising findings would reflect those that actually change the posterior distribution of entropy, which is by nature inclusive of all variables in the causal graph.

Preregistration of causal graphs is in fact a very feasible process because of our growing computational ability to depict and analyze graphs. I show practically how to do so with R code in the appendix for estimating uncertainty in entropy given a proposed sample size and research design with the method of Humphreys and Jacobs (2023). This formula could also be used for retrospectively assessing the cost and benefits of reproducing particular studies. The payoff of thinking about the entropy of causal systems is to allow for these relatively precise statements about the utility of research designs so that we can maximize our a posteriori learning without losing robustness.

Mechanisms and Causal Graphs

In this section I briefly consider an extension to the method that permits qualitative research to be included along with quantitative studies. By doing so, I show that the proposed method of entropy is flexible enough to permit modifications to the definition of causal graphs for non-quantitative research designs. To do so, we can consider a distribution of mechanisms in the causal graph V, which I denote as the set Ω. Each directed relation $P(x_i|pa_i)$ would have a corresponding distribution over mechanisms Ω, where $P(\Omega|x_i, pa_i)$ represents the joint distribution of all mechanisms for a given relation.

Specifying mechanisms in this way departs from some existing understanding of qualitative research like process-tracing as providing evidence over the existence of edges connecting variables in a causal graph (Humphreys and Jacobs 2023; Mahoney 2012). Incorporating mechanisms as separate entities permits mechanisms to be fundamentally distinct from variables, which captures the idea that mechanisms are root processes that connect variables as opposed to variables themselves (Waldner 2015) and for that reason may not be easily quantifiable (Collier 2011). While it is not the intent of this article to adjudicate between these competing theories of mechanisms, it is important to note that entropy is flexible enough to allow for such non-traditional understandings of causal graphs.

Because of causal graphs' Markovian property, we can simplify the expression of the joint probability of mechanisms to $P(\Omega|x_i)P(pa_i)$ because each variable in a causal graph is independent of other variables in the causal graph once we factor in its parents. In other words, we only need to consider the mechanisms of each node but not each preceding node. We can then consider the joint of both distributions:

$$H(P(v,\Omega)) = -\sum_{i=1}^{N} P(x_i|pa_i)P(\Omega|x_i)P(pa_i)\log P(x_i|pa_i)P(\Omega|x_i)P(pa_i) \tag{8}$$

If we consider a graph V that had identical distributions for the probabilities of causal

relations $P(v)$, but different distributions for mechanisms $P(\Omega')$, we could then define that a study that increased our type I understanding of mechanisms $P(\Omega')$ would be preferable iff:

$$H(P(v, \Omega)) > H(P(v, \Omega')) \tag{9}$$

In summary, this application of entropy to causal graphs reveals how the concept of statistical entropy can shed light on the difficult decisions that must be made when considering research designs. As I explicate in the following case studies, it is straightforward to calculate the amount of type I causal learning from plausible interventions on causal graphs using these formulae. Ultimately, the criterion of entropy suggests that we aim for maximizing the amount we can learn, i.e. causal knowledge, from an application of any of the paradigms. Even weakly causally-identified designs can contribute causal knowledge, and so long as causal graphs are preregistered, can allow for robust exploratory inference that aims at maximizing causal learning and minimizing the costs of expensive data collection.

While I do not include further examples of type II causal learning, its delineation here is to show the flexibility of this method of comparing research designs, including those designed to test for the credibility of prior work. The many formulations of entropy, including Kullback-Leibler divergence (Kullback and Leibler 1951), allow us to subsume equally as many research designs that go beyond exploratory inference. For example, this framework could be combined with meta-analysis to allow for entropy-weighted contributions from different research designs.

Case Study: COVID-19 Vaccines

In this case study, I show how even a relatively simple three-variable causal graph, along with relevant entropy calculations, can cover a range of possible studies, including experimental,

observational, and qualitative analyses. Importantly, all of the information necessary to understand the amount of causal learning can be inferred directly from the causal graph (although this is a post-hoc analysis as the studies I discuss did not have this causal graph preregistered). The aim of the case study is to show this method can be practically applied to actual research problems. In the appendix, I consider an additional case with a more complicated causal graph involving the relationship between oil and authoritarianism in the Arabian peninsula.

The outbreak of the COVID-19 pandemic offers an important test case for understanding how researchers employed research designs to understand and prevent COVID-19 infections. Because of the speed of the outbreak and the enormous scale of research efforts, there was relatively little time for traditional disciplinary norms to determine research designs. The SARS-CoV-2 virus did not care for disciplinary preferences, forcing researchers to employ whatever methods they had available to study the pandemic. The extreme pressures of this exogenous shock is the reason why I choose this particular area of research even though it is not a part of the social sciences proper. Under pressure, scientists went with their causal intuitions as opposed to relying on traditional disciplinary hierarchies, producing innovative work that built across modes of inference rather than relying on one mode at the expense of others. In addition, the massive levels of funding available from governments overcame one of the most common non-methodological factors in determining research designs, permitting issues of inference to become relatively more important.

Although there are many possible research questions, in this case study I focus on one crucial area: the development of vaccines. At first blush, it would seem that vaccines are a relatively straightforward exercise in terms of research design. After months of development, the drug companies Pfizer-BioNTech and Moderna released studies describing massive RCTs employing hundreds of thousands of volunteers over months. These studies provided concise and clear numbers concerning vaccine efficacy or the ratio of infected individuals in the

treatment group to the number infected in the control group (Polack et al. 2020; Baden et al. 2021). Because the control group never received a vaccine, the difference between the two groups could be directly attributed to the drug.

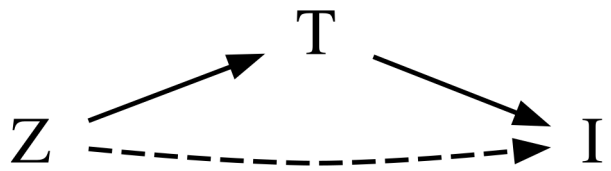

Figure 2: Directed Acyclic Causal Graph for Confounded Vaccine Uptake

In causal terms, this RCT solved a difficult yet well-known problem in studying COVID-19 infections: those who voluntarily participate in a COVID-19 vaccination study could be either more or less likely to be infected compared to those who would not want to volunteer for a vaccine. To give just one example, younger individuals showed less interest in vaccines compared to older individuals and were also much less likely to become severely ill. On the other hand, younger people may have been more likely than older people to contract a COVID-19 infection because they had less fear of serious illness or death. As a result, any naive comparison of a group of volunteers and the general population could end up conflating age differences with vaccine uptake (Hodgson et al. 2021; Baack et al. 2021). This causal identification problem is a straightforward example of confounding, as shown in Figure 2. Any variable which could explain both vaccine uptake and the incidence of COVID-19 would be a confounding variable, and without confidence that we can collect data on and measure all confounding variables, we may not be able to identify the direct relationship between the vaccination and efficacy.

As is well-known, the Pfizer-BioNTech and Moderna trials proved to be a paragon of RCT methods, showing remarkably strong effects of the vaccine on efficacy, above 90%. This type

of causal analysis was about as rigorous as possible for determining the precise nature of this bivariate relationship. At this point in the narrative, it would seem that RCTs had proven themselves as the gold standard: we had established that the vaccines worked, and now we could move forward with ending the pandemic. Indeed, such sentiments were common when the vaccines were introduced, leading to a relaxation of restrictions in the summer of 2021 (Bauer et al. 2021; Tregoning et al. 2021).

Fairly quickly, however, it became evident that the RCTs themselves were not sufficient to answer all the questions about vaccine efficacy. There were two main problems: first, people wanted to know how the vaccine performed in the population, which required attention to the confounding variables that the RCT successfully ignored (Hungerford and Cunliffe 2021), i.e., the role of the vaccine in the broader causal system. Second, the arrival of vaccine variants forced a re-evaluation of the vaccines' efficacy as RCT trials could not be run fast enough to keep up with new variants (Andrews et al. 2022). These issues required both observational and mechanism-based modes of inference, as I will explicate below.

Table 1: $Pr(I|T, Z = \text{Old})$

	I = Infected	I = Not Infected
T = Vaccine	.15	.85
T = No Vaccine	.85	.15

To describe these studies in terms of learning about causal systems, we can calculate the entropy reductions of these different interventions on a causal graph. For simplicity, I will take as my starting point the causal graph in Figure Figure 2, where the nodes are labeled as T for vaccine, one outcome, I for infection, and a confounder Z , which I will consider to be age. To simplify matters, I will treat each variable as having two discrete values.

To analyze the entropy of the causal process that produces I, we need to create the con-

ditional probability distribution given the two nodes that both directly affect I. Tables 1, 2, and 3 show conditional probability distributions for the three variables with plausible values. To analyze the relationships in the causal graph, we need to consider two conditional probabilities, $Pr(I|V, Z)$ and $Pr(T|Z)$, and one unconditional probability, $Pr(Z)$. Because the first conditional probability involves two conditioning variables V and Z, I separate this distribution of I into two separate tables for Young and Old subjects as can be seen in Tables 1 and 2.

Table 2: $Pr(I|T, Z = \text{Young})$

	I = Infected	I = Not Infected
T = Vaccine	.02	.98
T = No Vaccine	.98	.02

Table 3: $Pr(T|Z)$

	T = Vaccine	T = No Vaccine
Z = Young	.1	.9
Z = Old	.9	.1

I assume here that these are the true probabilities of treatment efficacy and the confounding effect of age on vaccine uptake. It is important to note, too, that age also affects vaccine efficacy, with the vaccine more efficacious among the young than the old, as studies have shown (Bell and Kutzler 2022). To calculate entropy, we will need to start with a prior distribution representing what we think these relationships could be, or what I will call the null graph. For simplicity, I will assume a uniform prior for the null graph, i.e., that all of the probabilities in the tables are equal to exactly 0.5. While not shown, I can calculate the entropy by considering the full joint distribution of the causal graph, which involves creating

a much larger table for all values of Z, T, and I, i.e. $P(I, Z, T) = P(I|T, Z)P(T|Z)P(Z)$. For reference, I also assume that the population is 25% young and 75% old for $P(Z)$.

Calculation of Shannon entropy[3] indicates that the null prior graph has an entropy of 209, while the true graph has an entropy of 124. These two figures give us the relative space within which we can plausibly learn about this outcome. If we end up with an entropy of less than 124, we will be over-confident, inferring causality to what are in fact random events. Respecting the lower bound reflects the principle of maximum entropy discussed earlier: we should not want to be more certain of conclusions than the underlying causal process permits.

In this case study and the one that follows, though, I will focus on type I causal learning. We can directly calculate the reduction in entropy for the experimental analysis of the vaccine by inserting values of the conditional probability distributions from the true graph for $P(I|T, Z)$ into the null graph. By adding in the true values for the conditional distribution, not just the average treatment effect, I assume that the treatment was high-powered enough to inform us about the true joint distribution of both the treatment and the potential confounder, age, as seemed to be true for most of the vaccine trials with tens of thousands of subjects enrolled. The entropy of this high-powered experimental analysis is 176, which is only 19% larger than the true entropy. As such, the experimental technique was clearly a powerful way of learning about this causal process.

However, the analysis left important information unanswered on the causal graph, in particular, what the relationship is between vaccine uptake and age. While the relative entropy would seem small, it is still an appreciable amount, and when vaccines were deployed to the population, it became a crucial factor for understanding the success of the vaccines (Hodgson et al. 2021). If sicker people were more likely to take the vaccine, then the measures

[3]R code showing how to reproduce all calculations in this case study will be included with the article upon publication.

of vaccine efficacy from the total population would understate the efficacy of the vaccine. Understanding how the vaccine interacted with population demographics required obtaining data about vaccine uptake in the "real world" (Chodick et al. 2021), especially to combat misinformation about the efficacy of the vaccine by anti-vaccination groups. For this reason, it is clear that even though the most important question about the vaccine was answered by an RCT, there was ample room for observational studies that collected data on the spread of the vaccine and relative rates of COVID-19 incidence in the population.

Ultimately, these observational studies were necessary to uncover the remaining entropy in the causal graph, equivalent to a 52% reduction in entropy. While this reduction was not as large as the reduction due to the experiment, it is important to note that this reduction could not be obtained from the experiment itself as it involved fixing the vaccine node V to a particular value, such as with Pearl's *do* operator. In this case, causal identification by fixing T to a specific value in an RCT prevented any analysis of the $P(T|Z)$ relationship because by definition it removed that causal arrow from the graph. For this reason, an observational analysis that established the $P(T|Z)$ relationship—varying vaccine interest by age—would likely be labeled as descriptive, not causal. However, this distinction is relatively arbitrary when we consider learning about the causal system as a whole.

Finally, it is important to note that mechanistic analysis also played an important role in determining the efficacy of vaccines in the pandemic. As mentioned earlier, the success of the vaccines waned depending on the mutations of the virus, and it was infeasible to keep running large RCTs for each variant. This is a kind of threat to inference that is rarely discussed, and could be described as "temporal validity" (Munger 2019). To address this problem, scholars examined whether the same mechanism underlying the vaccine's efficacy also occurred in the same way with variants, namely, the production of virus-neutralizing antibodies. These studies were not necessarily statistical in nature, involving close examination of relatively small numbers of petri dishes with new SARS-CoV-2 variants in blood that had vaccine-

induced antibodies (Yadav et al. 2021; Hoffmann et al. 2022). Furthermore, these studies could not be directly integrated into the causal graph examined above because they refer to factors that are not present on the graph itself, i.e., minuscule changes in antibody levels.

We can expand the analysis to incorporate the antibody mechanism if we give it two values, High and Low. With these two values in the set Ω, we can then calculate the entropy of the combined causal graph conditional on this mechanism for the relationship between vaccine V and infections I:

$$H(P(I, Z, T, \Omega)) = H(P(I|T, Z)P(\Omega|T)P(T|Z)Pr(Z)) \tag{10}$$

If we start with a null graph where the probability of Ω = High is 0.50, and the true value is 0.9, then we have null and true entropy values of 279 and 156 respectively. If we performed the experiment successfully but without learning about mechanisms, we would obtain an entropy value of 245 while an experiment that also involved learning about mechanisms would result in an entropy of 208, or 37% less. As can be seen, even with a single mechanism, relatively large reductions in entropy are possible by obtaining evidence of its true value. The reason for this large reduction is due to the fact that we know that the mechanism must be present for the causal story to hold about immunological response.

Of course, it could always be possible to look at antibodies as a mediator and expand the causal graph, allowing us to use mediation analysis to formally test for whether antibodies mediate the vaccine (Imai, Keele, and Tingley 2010; VanderWeele 2016). What is important, however, is that statistical tests were not necessary to make evidentiary statements about the presence or absence of the proposed mechanism. The proximity of observation and the logical necessity of antibody levels changing helped obtain causal inference about the performance of the vaccine against new variants even if there was no uncertainty interval or p-value attached. Statistical methods may not work well in this mode of inference because

the attention to micro-processes often entails serious limitations in data collection in favor of richly textured information (Collier, Brady, and Seawright 2010).

The point of this case study was not to argue that one mode of inference was superior to the other, but rather how each kind–experimental, observational, and mechanistic–played a valuable role in type I learning about a causal system that explained the level of infections at a given point in time. At different stages in the pandemic, each of these modes of inference helped determine the relative vaccine efficacy both against variants and in the population as a whole. While the RCT achieved the highest reduction in entropy of the causal system, other modes of inference had an important role to play. Once cost-benefit factors are included, such as the need to learn about efficacy against new variants for the purposes of booster shots, smaller overall changes in entropy can still be very important from a social welfare perspective.

In the appendix I include an additional case study examining research into the relationship between authoritarianism and the role of natural resources in rentier regimes. This case study employs a more sophisticated causal graph relative to the one previously described and likewise shows how different varieties of research designs, including experiments and process-tracing, can yield varying yet important amounts of learning about causal systems. With preregistration, these metrics could help make exploratory analyses of equal importance to more conventional falsificationist tests without privileging one or the other.

Conclusion

The principle of entropy provides one helpful framework by imagining the benefit of a study from the relative reduction or inflation in the entropy of a causal system. By doing so, we can change preregistration from a falsificationist standard to a more holistic and flexible form of preregistering our general state of knowledge of a causal process. This method can also be

integrated into meta-analysis studies, especially if studies are explicit about the causal graph they seek to understand. Ultimately, the aim of the framework is to create a level playing field for diverse research designs—including both exploratory and confirmatory studies—so that we do not under-invest in important lines of inquiry that might yield important information about causal systems even if they fail to achieve causal identification of pairs of variables.

Competing Interests

The author has no competing interests to declare.

Data and Code Availability Statement

All data presented in this article is simulated from statistical distributions. All calculations shown in this article and referenced in the appendix are available from a public Github repository: https://github.com/saudiwin/causality/.

Ethics and Inclusion Statement

This article solely makes use of data that is generated from simulated statistical distributions and as such there are no ethical issues to report in terms of human-derived data collection among marginalized populations.

References

Agassi, Joseph, and Institute of Philosophy, Russian Academy of Sciences. 2022. "Popper and His Popular Critics: Thomas Kuhn, Paul Feyerabend and Imre Lakatos: Appendix." Epistemology & Philosophy of Science 59 (4): 181–88. https://doi.org/10.5840/eps202259465.

Alrababa'h, Ala', Scott Williamson, Andrea Dillon, Jens Hainmueller, Dominik Hangartner, Michael Hotard, David D. Laitin, Duncan Lawrence, and Jeremy Weinstein. 2022. "Learning from Null Effects: A Bottom-Up Approach." Political Analysis, April, 1–9. https://doi.org/10.1017/pan.2021.51.

Andrews, Nick, Julia Stowe, Freja Kirsebom, Samuel Toffa, Tim Rickeard, Eileen Gallagher, Charlotte Gower, et al. 2022. "Covid-19 Vaccine Effectiveness Against the Omicron (b.1.1.529) Variant." New England Journal of Medicine 386 (16): 1532–46. https://doi.org/10.1056/NEJMoa2119451.

Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2021. Theory and Credibility. https://press.princeton.edu/books/paperback/9780691213828/theory-and-credibility.

Baack, Brittney N., Neetu Abad, David Yankey, Katherine E. Kahn, Hilda Razzaghi, Kathryn Brookmeyer, Jessica Kolis, Elisabeth Wilhelm, Kimberly H. Nguyen, and James A. Singleton. 2021. "COVID-19 Vaccination Coverage and Intent Among Adults Aged 18–39 Years — United States, March–may 2021." Morbidity and Mortality Weekly Report 70 (25): 928–33. https://doi.org/10.15585/mmwr.mm7025e2.

Baden, Lindsey R., Hana M. El Sahly, Brandon Essink, Karen Kotloff, Sharon Frey, Rick Novak, David Diemert, et al. 2021. "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine." New England Journal of Medicine 384 (5): 403–16. https://doi.org/10.1056/NEJMoa2035389.

Bak-Coleman, Joseph, and Berna Devezer. 2024. "Claims about Scientific Rigour Require

Rigour." Nature Human Behaviour, September, 1–2. https://doi.org/10.1038/s41562-024-01982-w.

Bauer, Simon, Sebastian Contreras, Jonas Dehning, Matthias Linden, Emil Iftekhar, Sebastian B. Mohr, Alvaro Olivera-Nappa, and Viola Priesemann. 2021. "Relaxing Restrictions at the Pace of Vaccination Increases Freedom and Guards Against Further COVID-19 Waves." PLOS Computational Biology 17 (9): e1009288. https://doi.org/10.1371/journal.pcbi.1009288.

Bell, Matthew R., and Michele A. Kutzler. 2022. "An Old Problem with New Solutions: Strategies to Improve Vaccine Efficacy in the Elderly." Advanced Drug Delivery Reviews 183 (April): 114175. https://doi.org/10.1016/j.addr.2022.114175.

Berinsky, Adam J., James N. Druckman, and Teppei Yamamoto. 2021. "Publication Biases in Replication Studies." Political Analysis 29 (3): 370–84. https://doi.org/10.1017/pan.2020.34.

Blackwell, Matthew. 2013. "A Framework for Dynamic Causal Inference in Political Science." American Journal of Political Science 57 (2): 504–20. https://doi.org/10.1111/j.1540-5907.2012.00626.x.

Blair, Graeme, Alexander Coppock, and Macartan Humphreys. 2023. Research Design in the Social Sciences. Princeton University Press. https://press.princeton.edu/books/paperback/9780691199573/research-design-in-the-social-sciences.

Brodeur, Abel, Nikolai M. Cook, Jonathan S. Hartley, and Anthony Heyes. 2024. "Do Preregistration and Preanalysis Plans Reduce p-Hacking and Publication Bias? Evidence from 15,992 Test Statistics and Suggestions for Improvement." Journal of Political Economy Microeconomics 2 (3): 527–61. https://doi.org/10.1086/730455.

Caldwell, Bruce J. 1984. "Some Problems with Falsificationism in Economics." Philosophy of the Social Sciences 14 (4): 489–95. https://doi.org/10.1177/004839318401400404.

Chodick, Gabriel, Lilac Tene, Tal Patalon, Sivan Gazit, Amir Ben Tov, Dani Cohen, and

Khitam Muhsen. 2021. "Assessment of Effectiveness of 1 Dose of BNT162b2 Vaccine for SARS-CoV-2 Infection 13 to 24 Days After Immunization." JAMA Network Open 4 (6): e2115985. https://doi.org/10.1001/jamanetworkopen.2021.15985.

Clarke, Kevin A., and David M. Primo. 2007. "Modernizing Political Science: A Model-Based Approach." Perspectives on Politics 5 (4): 741–53.

Collier, David. 2011. "Understanding Process Tracing." PS: Political Science & Politics 44 (4): 823–30. https://doi.org/10.1017/S1049096511001429.

Collier, David, Henry E. Brady, and Jason Seawright. 2010. "Toward an Alternative View of Methodology: Sources of Leverage in Causal Inference." In, edited by Henry E. Brady and David Collier. Lanham, MD: Rowman & Littlefield.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Control Trials." Social Science & Medicine 210: 2–21.

Dorie, Vincent, Masataka Harada, Nicole Bohme Carnegie, and Jennifer Hill. 2016. "A Flexible, Interpretable Framework for Assessing Sensitivity to Unmeasured Confounding." Statistics in Medicine 35 (20): 3453–70. https://doi.org/10.1002/sim.6973.

Flake, Jessica, and Eiko Fried. 2019. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." PsyArXiv. https://doi.org/10.31234/osf.io/hs7wm.

Frankenhuis, Willem E., Karthik Panchanathan, and Paul E. Smaldino. 2022. "Strategic Ambiguity in the Social Sciences." Social Psychological Bulletin, November. https://www.psycharchives.org/en/item/e5bb9192-80a4-4ae4-9cda-5d144008196e.

Gelman, Andrew. 2015. "Statistics and the Crisis of Replication." Significance.

Gelman, Andrew, and Eric Loken. 2013. "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "p-Hacking" and the Research Hypothesis Was Posited Ahead of Time." http://www.stat.columbia.edu/ gelman/research/unpublished/phacking.pdf.

Gerring, John. 2006. "Causation: A Unified Framework for the Social Sciences." Journal of Theoretical Politics 17 (2): 163–98.

———. 2012. "Mere Description." British Journal of Political Science 42 (4): 721–46. https://doi.org/10.1017/s0007123412000130.

Green, Donald P., and Alan S. Gerber. 2003. "Reclaiming the Experimental Tradition in Political Science." In, edited by Ira Katznelson and Helen V. Milner, 805–33.

Hausser, Jean, and Strimmer. 2021. Entropy: Estimation of Entropy, Mutual Information and Related Quantities. https://CRAN.R-project.org/package=entropy.

Hodgson, Susanne H, Kushal Mansatta, Garry Mallett, Victoria Harris, Katherine R W Emary, and Andrew J Pollard. 2021. "What Defines an Efficacious COVID-19 Vaccine? A Review of the Challenges Assessing the Clinical Efficacy of Vaccines Against SARS-CoV-2." The Lancet Infectious Diseases 21 (2): e26–35. https://doi.org/10.1016/S1473-3099(20)30773-8.

Hoffmann, Markus, Nadine Krüger, Sebastian Schulz, Anne Cossmann, Cheila Rocha, Amy Kempf, Inga Nehlmeier, et al. 2022. "The Omicron Variant Is Highly Resistant Against Antibody-Mediated Neutralization: Implications for Control of the COVID-19 Pandemic." Cell 185 (3): 447–456.e11. https://doi.org/10.1016/j.cell.2021.12.032.

Huang, Melody Y. 2024. "Sensitivity Analysis for the Generalization of Experimental Results." Journal of the Royal Statistical Society Series A: Statistics in Society, March, qnae012. https://doi.org/10.1093/jrsssa/qnae012.

Humphreys, Macartan, and Alan M. Jacobs. 2023. Integrating Inferences: Causal Models for Qualitative and Mixed-Method Research. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=0E3XEAAAQBAJ&oi=fnd&pg=PR9&dq=info:2K-gFfO67ZYJ:scholar.google.com&ots=cmUb1DyO6s&sig=OG0g6Cr3tsPGVMf01QNhKFmkgUE.

Hungerford, Daniel, and Nigel A. Cunliffe. 2021. "Real World Effectiveness of Covid-19

Vaccines." BMJ 374 (August): n2034. https://doi.org/10.1136/bmj.n2034.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." Psychological Methods 15 (4).

Jaynes, E. T. 2003. Probability Theory: The Logic of Science. Cambridge University Press.

Kocaoglu, Murat, Sanjay Shakkottai, Alexandros G. Dimakis, Constantine Caramanis, and Sriram Vishwanath. 2020. "Applications of Common Entropy for Causal Inference." In, 1751417525. NIPS'20. Red Hook, NY, USA: Curran Associates Inc.

Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." The Annals of Mathematical Statistics 22 (1): 79–86. https://doi.org/10.1214/aoms/1177729694.

Little, Andrew T., and Thomas B. Pepinsky. 2021. "Learning from Biased Research Designs." The Journal of Politics 83 (2): 602–16. https://doi.org/10.1086/710088.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." American Journal of Political Science 68 (1): 160–76. https://doi.org/10.1111/ajps.12723.

Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." Sociological Methods & Research 41 (4): 570–97.

Mukher jee, Debabrata, and Makarand V Ratnaparkhi. 1986. "On the Functional Relationship Between Entropy and Variance with Related Applications." Communications in Statistics - Theory and Methods 15 (1): 291311. https://doi.org/10.1080/03610928608829122.

Munger, Kevin. 2019. "The Limited Value of Non-Replicable Field Experiments in Contexts With Low Temporal Validity." Social Media + Society 5 (3): 2056305119859294. https://doi.org/10.1177/2056305119859294.

Pearl, Judea. 2000. Causality: Models, Reasoning, and Inference. Cambridge University Press.

Polack, Fernando P., Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra

Gurtman, Stephen Lockhart, John L. Perez, et al. 2020. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine." New England Journal of Medicine 383 (27): 2603–15. https://doi.org/10.1056/NEJMoa2034577.

Popper, Karl R. 1963. "Science as Falsification." Conjectures and Refutations 1 (1963): 3339. http://www.kokdemir.info/courses/psk515/docs/KarlPopper.pdf.

Protzko, John, Jon Krosnick, Leif Nelson, Brian A. Nosek, Jordan Axt, Matt Berent, Nicholas Buttrick, et al. 2024. "Retraction Note: High Replicability of Newly Discovered Social-Behavioural Findings Is Achievable." Nature Human Behaviour, September, 1–1. https://doi.org/10.1038/s41562-024-01997-3.

Schakenberg, Keith, Zion Little, and Jacob Montgomergy. 2024. "Preregistration and Strategic Private Experimentation." Working Paper.

Schneider, Jesper W., Nick Allum, Jens Peter Andersen, Michael Bang Petersen, Niels Mejlgaard, and Robert Zachariae. 2023. "Is Something Rotten in the State of Denmark? Cross-National Evidence for Widespread Involvement but Not Systematic Use of Questionable Research Practices Across All Fields of Research." https://doi.org/10.31222/osf.io/r6j3z.

Scoggins, Bermond, and Matthew P. Robertson. 2023. "Measuring Transparency in the Social Sciences: Political Science and International Relations." https://www.econstor.eu/handle/10419/268345.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." The Bell System Technical Journal 27 (3): 379–423.

Slough, Tara. 2023. "Phantom Counterfactuals." American Journal of Political Science 67 (1): 137–53. https://doi.org/10.1111/ajps.12715.

Spirling, Arthur, and Brandon M. Stewart. 2022. "What Good Is a Regression? Inference to the Best Explanation and the Practice of Political Science Research." Working Paper. https://github.com/ArthurSpirling/InferenceToTheBestExplanation/blob/

main/paper/Spirling_Stewart_WhatGoodIsARegression.pdf.

Swedberg, Richard. 2020. "Exploratory Research." The Production of Knowledge: Enhancing Progress in Social Science 2 (1): 1741. https://books.google.com/books?hl=en&lr=&id=vlTMDwAAQBAJ&oi=fnd&pg=PA17&dq=how+much+research+in+the+social+sciences+is+exploratory&ots=lUoGn5QhXs&sig=3iOR6vkvavwNkMwmNjMbPJMI0f0.

Tee, Phil, George Parisis, and Ian Wakeman. 2016. "NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium." In, 1049–54. https://doi.org/10.1109/NOMS.2016.7502959.

Tregoning, John S., Katie E. Flight, Sophie L. Higham, Ziyin Wang, and Benjamin F. Pierce. 2021. "Progress of the COVID-19 Vaccine Effort: Viruses, Vaccines and Variants Versus Efficacy, Effectiveness and Escape." Nature Reviews Immunology 21 (10): 626–36. https://doi.org/10.1038/s41577-021-00592-1.

VanderWeele, TJ. 2016. "Mediation Analysis: A Practitioner's Guide." Annual Review of Public Health 37. https://doi.org/10.1146/annurev-publhealth-032315-021402.

Waldner, David. 2015. "Process Tracing and Qualitative Causal Inference." Security Studies 24: 239–50.

Wieczorek, Aleksander, and Volker Roth. 2019. "Information Theoretic Causal Effect Quantification." Entropy 21 (10): 975. https://doi.org/10.3390/e21100975.

Yadav, Pragya D., Rima R. Sahay, Gajanan Sapkal, Dimpal Nyayanit, Anita M. Shete, Gururaj Deshpande, Deepak Y. Patil, et al. 2021. "Comparable Neutralization of SARS-CoV-2 Delta AY.1 and Delta in Individuals Sera Vaccinated with BBV152." https://doi.org/10.1101/2021.07.30.454511.