The Fundamental Problem of Interpretive Inference*

Jasmine English† and Richard A. Nielsen‡

August 22, 2024

Abstract

In 1986, Paul Holland articulated the fundamental problem of causal inference and helped inspire a causal inference revolution in the social sciences. In this paper, we argue that there is a fundamental problem of interpretive inference, analogous to the fundamental problem of causal inference. Social scientists often face the task of understanding the meanings humans give to aspects of reality, conceptualizing them, giving them appropriate place in theories, and measuring them for empirical investigation. The fundamental problem of interpretive inference is that we never observe meaning directly, but rather infer it indirectly. The problem of interpretive inference is significant for causal inference scholars because it interacts with the problem of causal inference. The counterfactual model of causation relies on the Stable Unit Treatment Value Assumption (SUTVA): the assumption that there be no hidden or conflated versions of treatment. When a representation of meaning is sufficiently incomplete and that meaning is a cause or consequence of a cause, this results in hidden or conflated variables. We urge greater recognition of this problem and suggest that more use of interpretive methodology would help positivists convincingly solve it. Researchers should adopt empirical, interpretive approaches to inferring meaning-making from data, rather than assuming that their naïve intuitions about meaning are sufficient for valid inference.

Word count: 10,652, all inclusive

* Thank you to Tariq Thachil, Nick Ackert, Devin Caughey, Aidan Milliff, Timothy Pachirat, Elizabeth Parker-Magyar, Anastasia Shesterinina, Eleanor Knott, and Eleanor Freund for comments, and to Jerik Cruz for comments and research assistance. We also thank the participants of the 2022 ECPR Joint Session on "Writing Politics" and the 2022 APSA annual conference for comments on the broader project of which this paper is part.

† Postdoctoral Fellow, Stanford University, jenglish@mit.edu

‡ Associate Professor, Massachusetts Institute of Technology, rnielsen@mit.edu

1 Introduction

In 1986, Paul Holland articulated the fundamental problem of causal inference (Holland 1986) and, in so doing, helped to inspire a causal inference revolution in the social sciences. The fundamental problem of causal inference is that we can never observe what would have happened counterfactually had treatment not occurred. In this paper, we argue that there is a fundamental problem of interpretive inference, analogous to the problem of causal inference. Social scientists often face the task of understanding the meanings humans give to aspects of reality, conceptualizing them, giving them appropriate place in theories, and measuring them for empirical investigation. The fundamental problem of interpretative inference is that we never observe meaning directly, but rather infer it indirectly. All communication of meaning is necessarily interpreted and thus relies on some representation, but we can never be certain that a representation of meaning is sufficiently accurate and complete for a given purpose. Greater recognition of this problem would improve applied research by encouraging researchers to adopt interpretive empirical approaches to inferring meaning-making from data, rather than assuming that their naïve intuitions about meaning are sufficient for valid inference.

The problem of interpretive inference is significant for causal inference scholars because it interacts with the problem of causal inference. The most widely adopted model of causation in the social sciences is a counterfactual model, often called the Neyman-Rubin-Holland Causal Model. This model relies on the Stable Unit Treatment Value Assumption (SUTVA) which assumes that there be no hidden or conflated versions of treatment, to avoid conflating the counterfactual outcomes under these conflated treatments. When a representation of meaning is sufficiently incomplete and that meaning is a cause or consequence of a cause, this can result violations of SUTVA. Scholars who take the problem of causal inference seriously should take the problem of interpretive inference equally seriously when it is relevant, because causal inferences are as weak as the weakest contributing inference.

Consider a treatment variable T that causes different responses in an outcome variable Y depending on the meaning that subjects ascribe to T, a situation that we call a *meaning-dependent mechanism*. For example, a researcher might use names the names "Emily" and "Latoya" to signal race in an experimental study (DeSante, 2013), relying on assumptions about the meanings

subjects ascribe to those names. If one subset of subjects does not interpret the name Latoya as meaning that someone is black, or they infer other socio-economic characteristics from the name, this can threaten causal inference through violation of an assumption that Dafoe, Zhang, and Caughey (2018) call "informational equivalence," which we recognize as a problem of interpretive inference. If a researcher incorrectly interprets what treatment means to their subjects, they may make incorrect inferences about their causal theories as a result. When a critic offers the challenge, "your treatment doesn't mean what you think it means," they are invoking the fundamental problem of interpretive inference.

Manifestations of the fundamental problem of interpretive inference have been identified already (Dafoe, Zhang, and Caughey 2018; Fong and Grimmer 2023; Pryzant et al 2021) and even identified as a "fundamental problem" (Egami et al 2022). Yet despite this important prior work, applied researchers are not sufficiently aware of the threats the problem of interpretive inference poses for their research, nor are they aware of prevailing best practices for interpretation in the social sciences. We see three impediments: the problem is not widely recognized or articulated through a unified framework, research practices in the causal inference tradition do not encourage explicit interpretive inference, and causal inference scholars are not fully informed of methods and standards for interpretive inference. We address each in turn.

We raise awareness of the fundamental problem of interpretive inference, show how existing problems noted in the methodological literature stem from this fundamental problem, and suggest that greater use of interpretive methods would help social scientists improve their causal inferences. By formulating interpretation as a fundamental problem, we emphasize the urgency of addressing interpretive inference explicitly in causal analysis.

Despite the importance of correctly interpreting human meaning-making for valid causal inference, researchers typically obscure the process of interpretive inference because it can conflict with other goals central to the positivist enterprise: objectivity, replicability, and credibility of causal inferences. Yet ignoring meaning-making undermines the credibility of causal inferences when meaning is causally important in the social world. In a discipline where the emerging norm is that causal claims are treated with substantial scrutiny and evaluated rigorously and transparently (Samii 2016), it is striking that meaning claims are often treated as

self-evidently true and are not typically evaluated as rigorously or transparently. This might not be a problem if interpretive claims were inconsequential to causal research, but they often are. When causal theories involve meaning-dependent mechanisms, credible causal inferences require clear and convincing interpretive inferences.

Experts in causal inference are rarely also expert in methods of interpretive inference. This lack of expertise is due, in part, to a deep disagreement about interpretive inference that divides social scientists into camps labeled positivist and interpretivist. These categories encompass a diversity of positions that are often taken for granted by researchers who work mainly in one camp. Positivism, strictly defined, has largely fallen out of favor in the philosophy of science, but it remains the methodological label approaches that are objective, replicable, and focused on the goals of causal inference and generalization (Marsh and Furlong 2002, 20; Lake 2013). Interpretivists, by contrast, favor approaches that are experiential, reflexive, and focused on the goal of subjective understanding. Because of this divide, scholars focused on causal inference are often blind to the challenges of credible interpretive inference, just as scholars focused on meaning-making are sometimes blind to the challenges of credible causal inference. Recent enthusiasm for mixing methods is tempered by calls from both sides not to mix too much across these traditions, lest interpretivism contaminate positivism, or vice versa (Ahmed and Sil 2012, Marsh and Furlong 2002).

We believe the division between methods for causal inference and interpretive inference must be bridged if problems of interpretive inference are to be more fully addressed in the context of causal research. Although causal inference scholars have independently developed some methods for interpretation, we argue that causal inference scholars put themselves at a disadvantage when they ignore the full repertoire of methods developed by interpretivists. We introduce several of these methods an discuss how them might improve interpretive inference by positivists. Regardless of the tradition from which they come, we argue that good interpretive inferences will generally have two properties: (1) they will explicitly consider multiple alternative interpretations with relevant data, and (2) they will describe the process of interpretive inference. We are hardly the first to advocate for tighter integration of interpretive and positivist approaches (Paluck 2010, Simmons and Rush Smith 2017, Thatchil 2018, Simmons and Rush Smith 2021), but

these proposals have tended to be limited one positivist approach (experiments, comparative case studies) have not yet been widely adopted.

2 Some Variations of the Fundamental Problem of Interpretive Inference

The fundamental problem of interpretive inference is more obvious in some settings than others, but it is widespread. Many already-recognized problems are variants of the fundamental problem of interpretive inference. Bringing these problems together under a single framework helps us see them more clearly and provides opportunities for solutions developed in one domain to be useful in others.

Ambiguity in the Meaning of Treatments is one instance. Experimenters try to study the effects of certain attributes on subject responses with experimental conditions that researchers think will induce certain meanings in a subject's mind. An example discussed by is the challenge of signaling race through names in experiments. In an important study, Desante (2013) showed that subjects respond differently to the "Black names" Keisha and Latoya (347, Table 1) than the names Laurie and Emily when evaluating deservingness for government benefits. When evaluating welfare forms in which the names are randomly varied, along with reports of whether the individual seeking welfare is "lazy" or "hardworking", respondents were more generous about allocating a constrained welfare budget to Laurie and Emily. However, subsequent reanalyses by Dafoe, Zhang, and Caughey (2018), Landgrave and Weller (2022), and Elder and Hayes (2023) suggest that the names also induce subjects to make inferences about other characteristics related to the outcome attitudes, such as socio-economic status. Whether this is a problem depends on the meaning of race that researcher intended to induce, which could include or exclude subject inferences about SES. But regardless, the experiment is ambiguous in the meaning of the name for subjects (presumably to avoid transparently communicating to subjects that the goal of the study is to make inferences about the effect of race), and thus subjects take the names to signify other meanings as well. A respondent who brings different background knowledge and thus ascribes a different meaning to the name might not receive the intended treatment at all.

The problem of ambiguous meanings in treatments extends to other shorthand signifiers that political scientists frequently deploy in experiments. In an influential study, Tomz and Weeks (2013) estimate the effect of democracy on the attitudes of residents of the US and UK about war by describing hypothetical potential adversaries in vignettes and varying whether the countries were described as democracies or not. Dafoe, Zhang, and Caughey (2018) note that manipulating the words "a democracy" or "not a democracy" in the vignettes also potentially changes subjects' assumptions about other characteristics of the hypothetical adversary. To subjects, the word "democracy" in the vignette seems to mean more than the institutions of government; it also means something about where a country is likely to be geographically located, its alliance portfolio, and its demographic composition. New research suggests that the demographic composition assumed by respondents is especially important. Rathbun, Parker, and Pomeroy (2024) show that "the term 'democracy' unwittingly primes presumptions of whiteness; respondents assume that nondemocracies are non-white. This implicit racialization, which varies across individuals, explains the reluctance of the U.S. public to support aggression against fellow democracies" (1).

Dafoe, Zhang, and Caughey (2018) term this problem the "assumption of informational equivalence:" an informational treatment manipulates only the meanings the researcher desires to manipulate, and no others. The solution, in this framework, is to anticipate and separate out the effects of the information and the "background" factors by accounting for the mediating effects of background factors explicitly (Acharya, Blackwell, and Sen 2018), as in Rathbun, Parker, and Pomeroy (2024). However, even this may not be adequate. Even if the "democracy" treatment only changed subject beliefs about government institutions (rather than racial demographics), it still does not follow that subjects have the same definitions of democracy in mind; for example, see Ridge (2023), on definitions of democracy in the Arab world. Democracy may be a "bundle of sticks" treatment (Sen and Wasow 2016), where invoking just the term does not uniformly manipulate the bundle in the same way for all experimental subjects. This problem is not solved by signaling meaning in some other way, because other representations of meaning are also limited. Benstead, Jamal, and Lust (2015) innovate by signaling religiosity of hypothetical electoral candidates through photographs showing variation in religious clothing rather than

textual assertions about religiosity. However, they acknowledge that "the veil does not have a single meaning" (78); the religious meaning intended by the researchers could have been misinterpreted by subjects in a variety of ways. Moreover, the photographs convey information other than religion, though the researchers attempt to control this (78).

Fong and Grimmer (2023) treat the problem of interpretive inference in a different way. They also consider a situation where respondents are exposed to text and must make meaning of it, but in their formulation, the confounding comes from latent treatments in the text, rather than background beliefs. For example, an experiment might estimate the effect of negative information in campaign advertising on the decision to vote (Fong and Grimmer 2023, 374), but the advertisement also provides information about where to vote, that is not measured by the analyst. When multiple treatments are present, "the treatment of interest may be *aliased* by other measured or unmeasured latent treatments," resulting in treatment effects that capture both the intended and unintended treatments. Combining these aliased treatments as part of "treatment bundle" is unsatisfying because this disconnects the empirical result from the treatment that relates to theory. Yet manipulating treatments independently may be impossible; there may be no way to provide text that manipulates only a single treatment. The problem is compounded outside of experimental research, when scholars seek to make inferences about the effects of text in which embedded treatments are determined non-randomly by authors.

Pryzant et al (2021) focus on the problem of estimating the effects of "linguistic properties," in which the potential intervention is telling a writer to write with a certain property or not. For example, they pose the problem of estimating the effect of whether the writer of a customer complaint intends to be polite or not. Pryzant et al "we imagine intervening on writers and telling them to use the linguistic property… because the hypothetical intervention is well-defined" (3). While we agree that the treatment is well-defined in one sense, it is poorly defined in an obvious way that invites multiple interpretations: what is "polite"? The solution of Pryzant et al is to make inference about author intent from the reader side: "Readers use the text W to perceive a value for the property of interest…as well as other properties…then produce the outcome Y based on these perceived values" (3). This makes the role of interpretation explicit at

two stages: from a hypothetical intervention where the treated writer must interpret writing politely and the reader must subsequently interpret politeness from the text. Identification is possible in this setting with assumptions that we find to be strong: "if we assume that readers correctly perceive the writer's intent, the effect…in terms of observed variables, is equivalent to the effect that we want" (3). Their applications are to situations where we can generally only learn proxy labels of a linguistic property from the text, rather than accessing direct indicators of either the author's intent or reader's perception. This is practical for observational data online, but adds yet another layer of interpretive inference: that the proxy labels learned for the linguistic property are interpreted correctly by the researcher as well. While the results in this paper push causal inference with text forward, it is with strong assumptions about interpretation that we argue would benefit from direct data collection that informs the interpretive inference.

Our discussion has focused on problems that arise from the ambiguity of treatments, but ambiguous meanings of outcomes, confounders, moderators, and instruments all potentially affect what researchers should conclude from their analyses.

3 Interpretation of Meaning in Causal Inference

Because humans interpret the physical and social world to have meanings, the challenges of observing and accounting for these meanings is a necessary part of social science. The fundamental problem is that meanings cannot be observed; they can only be indirectly represented. We represent meanings to each other using symbolic systems. To understand the meaning of a symbol is to deploy it correctly, in the right context and for the right reasons. These systems function remarkably well but are nonetheless representations and thus susceptible to misinterpretation.

This is not a new claim; we are merely restating an old and widely recognized problem. Charles Taylor explains the problem in his 1971 essay on interpretation in Political Science: "meaning has an essential place in the characterization of human behavior," yet establishing criteria for adequate interpretation of meaning presents deep challenges because there are no objective criteria for declaring an interpretation correct. "A successful interpretation is one which makes clear the meaning originally present in a confused, fragmentary, cloudy form. But how

does one know that this interpretation is correct?" The failure to fully convince someone else of an interpretation raises a fundamental challenge to confidence in our own interpretations. "if I am this ill-equipped to convince a stubborn interlocutor, how can I convince myself? How can I be sure? Maybe my intuitions are wrong or distorted, maybe I am locked into a circle of illusion." (51).

Taylor was critical of mainstream Political Science for, in his view, attempting to sidestep the challenge of interpretation by resorting to "brute data", observations with "no element in it of reading or interpretation."

> It can be argued then, that mainstream social science is kept within certain limits by its categorial principles which are rooted in the traditional epistemology of empiricism; and secondly, that these restrictions are a severe handicap and prevent us from coming to grips with important problems of our day which should be the object of political science. We need to go beyond bounds of a science based on verification to one which would study the inter-subjective and common meanings embedded in social reality. But this science would …not be founded on brute data; its most primitive data would be readings of meanings (94)

Perhaps it was true when Taylor wrote, that social scientists followed strictly an epistemology of empiricism and eschewed investigations of meaning. But now, many political scientists are interested in phenomena where meaning plays a central role.

We agree with Taylor that when social scientists measure meanings, the primitive data are necessarily interpretations – readings of meanings, in Taylor's phrasing – rather than "brute data." These interpretations are symbolic models that attempt to convey a meaning. The meaning is separate from the symbolic model and multiple symbolic models can be used to convey the meaning from one human mind to another. For example, in this paper, we try to keep our meaning consistent but we elocute around it in many ways. These different ways are all necessarily incomplete attempts to convey our meaning; we hope that combining them makes our meaning clear.

If representations are symbolic models, then following Clarke and Primo (2013), these models are objects that are simplifications of reality. Clarke and Primo hold that objects cannot

have a truth value, so a model cannot be said to be false. While we find this view useful, we differ somewhat because linguistic models might be said to have a truth value (Godfrey-Smith), which would allow us to say that a particular representation is a "misinterpretation" of meaning. We take this claim to mean that in the process of transferring a meaning from one mind to another via a set of representations, the resulting meaning is importantly different from the original meaning. We might blame the representation offered or the interpretation of the representation by the recipient. But retaining the ability to call something a misinterpretation in some settings is important.[1]

Because "uncertainty [about interpretation] is an ineradicable part of our epistemological predicament," (Taylor 1971, 51), we should account for this uncertainty when we make causal inferences about the causes and consequences of meanings.

Formally, consider that a treatment Z interacts with a meaning *M* assigned by the subject to that treatment to produce an outcome in the presence of potential confounders, both observed and unobserved. This corresponds to the situations above in which an experimenter wants to know the causal effect of changing the information available to a subject and that information is ambiguous. To make causal inference without investigation of *M*, it must be the case that *M* varies deterministically with Z. In the terms of an example above where names are used to signal the race of a character in an experimental vignette, if Z perfectly predicts *M*, then random assignment of Z will allow us to get an unbiased estimate of the effect of Z on Y.

If, however, *M* can take on two values that to do not covary perfectly with Z, then failing to observe *M* may lead to biased estimates. For example, if some respondents see the name Latoya and it signals only race, while others see the name signaling both race and socio-economic status, and there are potential confounders that influence why some make one interpretation and others make the other, then we have a problem. If the goal is to make inference about *M*, then we can think of this as a non-compliance problem. The researcher wanted to induce *M*=*m* but in some subjects induced *M*=*m**. Inference about cases where Z causes M=m are confounded by unwitting

[1] Allina-Pisano (2009, 63) describes an example of a potential misinterpretation and how "the type of knowledge acquired through ethnography allows the researcher to discern more intelligently the material condition of such a village street, and to avoid a misinterpretation that might so easily be produced through the use of other qualitative methods."

inclusion of the cases where M=m*. The factors that influence whether subjects have *M=m* or *M=m** are not randomize, so random assignment of Z no longer justifies a causal interpretation of the effect of *M* on Y.

Dafoe et al represent the problem as one of "background beliefs" (B) that are inadvertently activated by a treatment Z. When treatment causes changes in background beliefs, and background beliefs affect beliefs about the key variable of interest, then the effect of D on Y is confounded unless B is controlled.

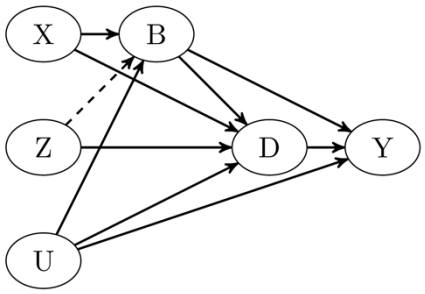

Figure 1: Dafoe et al representation of the "information equivalence" problem. Z is the manipulation, X is other vignette scenario details, B is background beliefs, D is beliefs about the causal factor of interest, U is unobservables, Y is the outcome. If the dashed arrow from Z to B is absent, then Z instruments D.

In the Dafoe et al framework, background beliefs are conceptually distinct from beliefs about the causal variable of interest. This privileges the researcher's interpretation of the key variable by representing it with D, and treats alternative interpretations by subjects as separable into the same interpretation D and additional beliefs B that must be measured and accounted for. For example, if a vignette manipulates whether a country is "democratic," then this approach assumes that all respondents understand some common concept of democracy, but that they might also have background beliefs that are accidentally manipulated by treatment, such as the geography or demographics of the country. But what if some subjects do not understand democracy in any overlapping way? Imagine that researches assume that democracy entails free elections, but some subjects do not share this interpretation. The fact that democracy means something different to these subjects is hardly a "background belief" because it is the belief of interest. But heterogeneity in interpretations of democracy may be related to X and U, meaning

that although Z is randomly assigned, the realized level of D is correlated with X and U, confounding the estimate of D on Y.

Fong and Grimmer address the problem differently, focusing on effects of the text itself, rather than background beliefs. They consider cases in which exposure to text X contains both intended treatments Z which are measured, and unmeasured latent treatments B which causally affect the outcome Y.

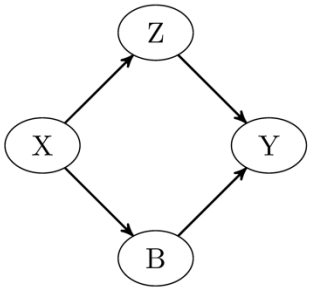

Figure 2: Fong and Grimmer representation of "latent treatments" from text. X is the text, Z is the intended treatment, B is the unmeasured latent treatment, Y is the outcome jointly caused by Z and B.

Fong and Grimmer represent the translation from texts X to the observed treatment Z using a codebook function g, such that $g(X) = Z$.[2] There is an analogous codebook function h for the latent treatments, but it is unmeasured and thus unknown. Fong and Grimmer assume the codebook function g to be fixed. If g is fixed, then when it is applied to X, it directly converts X into latent treatments without ambiguity. This sidesteps the core problem of interpretive inference by placing responsibility for the codebook function onto the researcher, rather than showing how to constructed a useful g. If the researcher knows the right codebook function, then accounting for unmeasured latent treatments means identifying additional codebooks h that correspond and making those explicit. In our argument, however, the researcher does not know the codebook function and we expect that there is heterogeneity in the codebook function among respondents that matters. Consider a campaign ad that tells viewers that "a jury determined that Trump paid hush money to a porn star." A researcher might reasonably create a codebook that classifies this statement as "negative" or devise an experiment in which this statement was the "negative advertising" condition. Many respondents might interpret this statement as negative.

[2] They subscript to i throughout, so I need to update our notation.

Yet some respondents might interpret this statement as positive information about Donald Trump's sexual prowess, social status, and economic status. In Dafoe et al's framework, we might think of the text as a negative treatment that also shifts other background beliefs about trump that are positive. In Fong and Grimmer's framework, we might see this text as embedding both a negative treatment (Trump did something illegal) with several latent positive treatments (Trump has money and access to sex workers). Both of these may be useful models, but we find that they do not directly address the issue that the statement simply means different things to different people, and that for some, the sentence has no negative meaning. Thus, while both Dafoe et al and Fong and Grimmer assume that the treatment is delivered and the problem is background beliefs or aliasing treatments, our approach highlights that the same text delivers different treatments if we define treatment as the meaning that respondents make of the text.

4 How to Improve Interpretive Inference

Positivists are already doing interpretive inference each and every time they encounter an interpretive problem and we see no escape from it. Research that involves meaning cannot avoid the problem of interpretation, but many researchers try. Our aim is to encourage positivists to do interpretive inference explicitly and well.

Our claim isn't that positivists should switch to being expert practitioners of interpretive methods. Instead, even passing familiarity with interpretive approaches can greatly improve positivist research. Gate-keepers might protest that these methods can be misused, and that without becoming card-carrying interpretivists, researchers risk making mistakes. Of course they do. But we believe that on balance, the problem is very much in the other direction: positivists are already making interpretive inferences all the time and familiarity with more tools will mostly help.

One important use of an interpretation is to help facilitate measurement of a variable that approximates the meaning that individuals make. If this is the use, then there can be misinterpretations, and thus our models can be right or wrong in an important sense. For this type of interpretive inference, solutions that look like validation are likely to result in good inferences. The general procedure is to develop a representation of meaning from one procedure

and then check it against representations developed from alternative procedures to see if the representations share commonalities or have differences that seem important to causal inference. We could theoretically start with any procedures, but to save time, analysts should start with procedures that seem like promising ways of representing meaning, either because we have reasons to believe so from first principles, or they have a track record of helping in the past. This task is analogous to the tasks of measurement or supervised learning – there is a target in the real world, and although we do not have direct access to the target, we can rely on indications of how close we are getting to discard misinterpretations (see, for example, Rasmussen et al 2024).

Another use for interpretive inference is to conceptualize or reconceptualize phenomena in the world. This often involves classifying observations; asking the question "what is this a case of?" Unlike interpretative inference where the target is to faithfully represent the key elements of meaning in someone else's mind, this type of interpretive inference does not have to correspond to any other person's interpretation (though it might). The criteria here is whether a new interpretation is useful. Does it reveal new aspects of a phenomenon that were previously hidden? This is analogous to the task of discover or unsupervised learning.

As with many other methods, the purpose of interpretive methods is to enhance to parts of the inferential task that humans are good at and compensate for parts of the task that humans are bad at. Humans are remarkable at making meanings, so we typically do not need tools to tell us that something might have a meaning. However, we tend to overinterpret meaning in some contexts (finding patterns in random dots) and we are more committed to our understandings of meaning that we out to be. This means that we are bad at considering a wide range of possible meanings, we are poor at articulating where we made meaning from, and we are biased towards interpretations that fit with what we already think we know.

We argue that interpretive inferences in positive work should meet the following standards. Researchers should 1) consider multiple possible meanings; 2) use data to inform their inferences about which possible meanings are relevant for research design and data analysis; and 3) transparently report their inferential process. In a survey of articles, we find that even when treatments invoke meaning from text and symbols, these three criteria are rarely met.

We coded how authors justified interpretations of informational variables in a sample of 51 articles from the APSR and AJPS coded by Fong and Grimmer in their analysis of latent treatments.[3] All of these articles had at least one important variable that could be interpreted in multiple ways by respondents. Sixteen studies had no acknowledgement of alternative possible interpretations at all. In 31 out of 51 (60%), the subjects' interpretation was assumed to correspond to the researchers' with relatively little justification. Only one, or maybe two studies showed reflexivity in their interpretations. Only one, Thachil (2017) matched all three criteria. Other papers that meet some of these criteria typically do so through in the language of "robustness." The tradition of "robustness checks" in quantitative research has many commendable qualities, but placing alternative potential interpretations as secondary analysis that appeases skeptical reviewers gives short shrift to the fundamental problem of interpretation. We conclude from this literature survey that the most common approach researchers take to interpretive inference is to apply their own interpretations, and consider alternatives only when forced to by critics or the review process. In some instances, this is justifiable because a meaning is commonly held by the researcher and all individuals in the research. In others, it is highly debatable that there is a single shared meaning, and that the researcher's own interpretation of meaning is the same.

A set of meaning-related tasks are common to many research projects: conceptualization, signifying meaning carefully (manipulation), inferring how meaning was understood (measurement), conditioning on meaning (conditioning). A set of common strategies have developed for accomplishing these tasks: "casing a study," ordinary language analysis, and close observation with an "ethnographic sensibility." Researchers can mix and match these strategies for their specific meaning-making task, and we expect that greater recognition of the problem will lead to new strategies or highlight the usefulness of strategies we have overlooked. We highlight how for some of these tasks, practices proposed by quantitative and qualitative methodologists are similar to each other, and thus complementary rather than competing. A

[3] Thanks to Jerik Cruz for excellent research assistance to code these data!

common theme in these approaches is attention to considering or collecting data that researchers might otherwise disregard.

Considering Alternative Interpretations

Considering alternative meanings is a foundational task for interpretive inference. Researchers cannot credibly claim to have the most appropriate interpretation of meaning-making if they have not explicitly considered any alternatives. But we cannot consider what we haven't thought of. We need methods that force alternative interpretations into our view for consideration.

Unsupervised methods for clustering texts and images offer a quantitative approach for generating possible interpretations (cite Consilience). Topic models are often used by researchers for this purpose – they force a researcher to contend with clusters of correlated words delivered to them by the model that they might have overlooked from merely reading (some of) the documents.

Interpretivists approach the problem of generating alternative interpretations for consideration with a different suite of methods. Soss (2021) describes the interpretive process of "casing a study": In this approach, a researcher problematizes a concept by asking questions such as "what could this phenomenon be a case of?" Trying on different categories for a phenomenon prioritizes consideration of multiple alternative interpretations. Our recommendation is that researchers making interpretive inference as part of positivist inference should report possible "casings" of key phenomena and concept. Practically, this could entail a statement or list, such as "an advertisement informing voters of Donald Trump's criminal conviction could be a case of a negative campaign ad, but also a case of an informative ad." Statements like this would inform readers which possible meanings of a phenomenon researchers have considered. This mirrors best practices in case selection, where specifying which cases were not chosen and why is important for justifying inference from "most similar" case study design (Nielsen 2016).

To "case a study," scholars rely on a number of tools for considering what the phenomenon they observe and manipulate might be cases of, including some of the tools we discuss below. But the most important thing is to unsettle the meanings that seem obvious to

researchers and give open-minded consideration to alternative meanings; as Soss says, to intentionally stretch concepts beyond their typical use, such as when he considers claiming welfare benefits to be a form of political participation.

Asking "what is this a case of?" can also matter causal inference down the road because understanding what something is a case of can determine what the comparison cases should be for causal inference, what the relevant controls should be, what the important causal variables are likely to be, and even what outcomes are important

Supporting interpretations with relevant data

Otherwise, interpretation is entirely based on researcher background, assumptions, prior knowledge. These may be accurate or inaccurate, but without incorporating relevant data into the judgment, there is no chance for information about the specific situation to correct a misinterpretation. There is no single type of data that is always best for interpretive inference. It could be quant or qual, experiential, etc. Interpretation of a piece of evidence depends on the context, and diverse pieces of evidence can matter.

To infer meaning, interpretivists often use data about the words and symbols people use to describe what phenomena mean to them. Ordinary language analysis is an interpretive approach for uncovering the meaning of words in everyday talk (Schaffer 1998; 2006). In this approach, the meaning of a term is strongly related to how it is used by people; to understand a meaning is to use a term appropriately in context. A related approach, called the *semiotic-practical approach* by Wedeen (2002), encourages attention to what language and other symbols do politically. These approaches to language and symbols can help positivists understand the meaning-making of political actors and justify some interpretations over others. We recommend that researchers provide evidence from ordinary language analysis that supports their inferences about meaning-making. This can include qualitative ordinary language analysis from interviews or close reading of text. It can include quantitative ordinary language analysis from word embeddings, topic models, or other quantitative summaries of text that summarize how words are used in context (see Kozlowski, Taddy, and Evans, (2019) for an example investigating meanings of social class).

We draw here on *Elucidating Social Science Concepts* (Shaffer 2006). For Schaffer (2006), "to elucidate concepts is to investigate the situated uses of their corresponding words." (89) The cover of the book illustrates with the word "power" in black and then lists it in a variety of contexts: "power to," "power over," and on down through "power lunch" before arriving back at "power" on its own. It is a visual journey in ordinary language analysis. Schaffer's approach is not purely introspective, however. "To investigate grammar ethnographically is to rely on interviewing or textual analysis rather than introspection as the main source of insight." (Shaffer p44).

Schaffer endorses qualitative approaches to this investigation. Schaffer uses the term "elucidating concepts" as a shorthand "for the self-conscious application of interviewing techniques inspired by ordinary language philosophy" (Schaffer 2006, 151). The purpose is to engage the interviewee in a conversation, and within that conversation, to investigate the meanings of words of interest with targeted questioning strategies. "Among these strategies are judgment questions that require the interviewee to express opinions that lay bare the standards implicit in a term; example prompts that invite the interviewee to recall or imagine concrete instantiations of a term; and internal-logic questions that provide the interviewee an opportunity to reflect on how various ways he or she is using a term might be connected" (Schaffer 2014, 308). Also see Schaffer (2006, 154-158) on how to conduct an ordinary language interview. Yet despite Schaffer's focus on interviewing, we argue that quantitative approaches to investigating the situated uses of words, such as word embeddings, are equally useful. Schaffer recognizes that searching through large databases forces analysts to confront uses of words they didn't expect (49) and endorses using the Google n-grams viewer to understand word birth and death (69).

As an example of elucidation, Schaffer considers the problem of defining "family" for political study (22). Sartori's definition of "legitimate heterosexual for rearing children" misses a lot of meanings people give to family, and these in turn both generate social science questions (how do people "do family"?) and could keep positivists from misunderstanding their own results (decline of the "family", channels of influence on voters in a study of how family influences votes). An ordinary language analysis approach investigates how "family" is used, rather than how family is "defined," and thus helps researchers avoid that trap of thinking

everyone shares their (dictionary) definition of a particular term. "Elucidation, in this regard, should be of interest not only to interpretivists but to self-aware, morally responsible positivists as well, for it reveals the particular limits and dangers of reconstructing a concept in this or that way" (Schaffer 2006, 22).

Positivists may find ordinary language interviews useful as they interpret actors' descriptions of their interpretations (i.e., as they interpret actors *in their own words*). Ordinary language interviews reduce the risk of erroneous interpretations due to misunderstandings of words in particular contexts—a significant risk, we think, given the ways in which even American English can vary by class, race, gender, profession, ideology, and sexual orientation. Ordinary language interviews, then, offer a systematic approach to assessing and checking what actors *actually mean*. For this reason, we think this interview strategy represents another possible solution to the interpretation problem, and one that can assist with interpretation in positivist research in ways that (a) reduce the possibility of incorrect interpretations and (b) guide future researchers. What might this look like in practice? One approach would be for researchers to incorporate ordinary language interviews into their fieldwork to guide ongoing data collection and emerging theoretical insights (i.e., as a way to answer the question, "am I getting it right"?). Alternatively, researchers could use these interviews *after* their analyses, to verify that they have understood the implications of particular words more or less correctly.

Schaffer (1998) demonstrates the value of ordinary language interviews in *Democracy in Translation,* which explores the differences between elite and ordinary citizens' understandings of what democracy means. Schaffer finds that Senegalese elites tend to invoke the word democracy in ways similar to the usage of many political scientists (i.e., a democratic system is one in which elections are contested and outcomes uncertain). By contrast, lower-class, less-educated Senegalese use the Wolof equivalent, *demokaraasi*, to mean "equality" or the attainment of "collective economic security via mutuality" (1998, 85). Schaffer shows that these linguistic distinctions have observable implications for behavior and political outcomes: *demokaraasi* has

effects on "the intentions and actions of voters and nonvoters" in ways that shape "the accountability of public officials and the exercise of power in the country."[4]

Beyond ordinary language analysis, Interpretivists attend to symbols and symbolism in ways that we think would benefit positivists. There are many ways to do so, but Wedeen (2002) offers an overarching approach for structuring this attention, which she calls the *semiotic-practical* approach. As outlined in Wedeen (2002), "A semiotic-practical approach investigates what language and other symbolic systems do—how they are inscribed in concrete actions and how they operate to produce observable political effects." In other words, a semiotic-practical approach investigates how language and symbols (signs, advertisements, speeches, official imagery) shape practices and behavior (e.g., political compliance, political discussion, political support). As Wedeen (2002, 723) puts it: "semiotic practices produce material effects, the observable implications of which are so important for positivist social science." Although Weeden focuses on interpretive, qualitative research when advocating for the semiotic-practical approach, we find this definition broad enough to include any causal analysis of symbolic systems on social outcomes.

Positivists may find the treatment of causal inference in Wedeen (2002) unsatisfying, but her approach calls attention to data that positivists sometimes ignore at their peril. Use of a semiotic-practical approach helps positivists by pushing them to identify the semiotic practices relevant to the given phenomenon and explore how such semiotic practices work. Consider, for instance, a study of the determinants of patriotism in the United States.

> "We might select pledging allegiance to the flag as one semiotic practice in the range a scholar investigates in studying patriotism in the United States. We would want to do more than analyze its content or infer symbolic patriotism from its family resemblance to flag ceremonies elsewhere. We would inquire into the effects of the relationship between pledging allegiance and patriotism, a task that calls for a number of additional studies. We might research the history of pledging allegiance, including the mechanisms by which the

[4] *Demokaraasi* plays important roles in "legitimating governmental coalitions, in making public decision making less participatory, in maintaining people's engagement with the state, in contributing to the pluralization of power, and in perpetuating the privileged position of the ruling party" (Schaffer 1998).

> ritual was enforced over time. We might observe the practice ethnographically in areas selected for their varying regional, ethnic, racial, class, and political affiliations. We might conduct open-ended interviews and surveys, asking a wide variety of people about the meanings they attribute to pledging allegiance and the effects it produces in them: Was the flag salute mind-numbing, uplifting, apathy-inducing, or irrelevant? Finally, we might collect transgressive materials, such as evidence from court cases and protest movements, as well as source materials from "popular culture" media, such as newspaper reports, films, jokes, cartoons, and songs, that may offer alternative ways of seeing the pledge of allegiance. Such an analysis would allow us to discern whether the pledge of allegiance could be a banal, routinized practice, an activity invested with and productive of patriotism, or both." (Wedeen 2002)

A semiotic-practical approach encourages a way of looking at the world that requires an account of how and why *meanings* generate action. It encourages attention to more varied forms of evidence than standard positivist approaches (e.g., a phenomenon's historic emergence, participant observation of the phenomenon, open-ended interviews about the phenomenon, and its representation in "newspaper reports, films, jokes, cartoons, and songs"). In Syria, Wedeen (1999) shows how the speeches, slogans, and imagery characteristic of the leader's "cult of personality" operated to enforce obedience, induce complicity, and set guidelines for public speech and behavior. While Wedeen's research is avowedly interpretivist, attention to symbols in context has proven useful in positivist work with interpretive inference. For example, Corstange (2012) pairs a traditional survey of attitudes in Lebanon with untraditional observations of publicly displayed religious iconography outside of respondents' residences, interpreting these symbols to understand the religious identities of Lebanese citizens. Non-linguistic symbols are potential data for understanding meaning-making.

Making Interpretive Inference Transparent

A major point of difference between positivists and interpretivists is terminology and philosophy around research transparency. Positivists will often seek to make work reproduceable and replicable, meaning that other researchers have access to all the data and

analysis instructions necessary to exactly reproduce a claimed result, and that analysts who investigate a finding with new data can obtain equivalent results. Interpretivists are philosophically inclined to see research findings as inherently subjective, so that reproduction and replication are impossible; there is no "objective" observation apart from each researcher. So interpretivists may prefer to hold research to a standard of transparency --- that data and inferences are carefully described --- without an attempt to make the "raw" data fully available or the inference algorithmic.

It is often more difficult to provide replication data and "code" for interpretive inference, but a transparent description of how the researcher made the inference and (possibly) supporting evidence that they made the inference the way they said they did should become standard practice. This does not necessarily need to increase the length of journal articles, in the same way that replication code doesn't need to increase the length of journal articles. It does increase the amount of public-facing material that researchers should make available, which increases the workload for researchers. Yet if transparency about causal inference is deemed an essential use of researcher time, then transparency about interpretive inference that affects the credibility of causal inference should also be an essential use of research time.

Reflexivity is a key feature of how interpretivists make inference explicit. A reflexive researcher notes their perception of how their particular mind, body, and social position might be affecting their inference. They often describe how they presented themselves to research subjects, how they believe they were perceived, and what they think respondents were willing or unwilling to tell them. For example, Cramer (2016) records how her affiliation with the University of Wisconsin affected her conversations with rural residents of the state.

Interpretivists also often recommend an "ethnographic sensibility." While this term has multiple potential meanings, the dominant one is that researchers bring an ethnographic mindset to their research, even when it is not data collection through participant observation during a long stay at a field site. This can mean attention to non-traditional data, to the "meta-data" surrounding data collection, and to the embodied experience of working with data. Lee Ann Fujii (2015) provides several examples of "accidental ethnography"—the unplanned moments during research (overheard stories, for instance, or observations of everyday scenes) that deepen the

researcher's understanding of the research context, and lead to discoveries that determine the course of the project. Schaffer's elucidation encourages researchers to attend to what words make them feel: "elucidation can…be enriching for reasons related to the feelings of awe, joy, sorrow, or rage that it awakens." (92). While positivist researchers are generally trained to keep their emotions out of research, interpretivism draws on researcher emotional responses as potential data. Even short stints of participant observation can illuminate for researchers how performing an action might feel to those who perform it, and this can transform what researcher concludes. Describing these moments of reflection, of non-traditional data collection, and of embodied insight will strengthen interpretive inferences in positivist work by bringing readers along for the inferential journey, even if it is hard to fully systematize.

Opportunities and Pitfalls

The foremost benefit to positivist scholars who consider applying traditionally interpretive methods in their efforts to address the problem of interpretive inference is *fidelity*: that our theoretical and empirical models usefully resemble the parts of the world we wish to understand. Without close attention to how their research subjects' meaning-making, researchers may measure meaning in ways are not faithful in important ways, and this can result in low-fidelity causal models as well.

Interpretive methods can also encourage researchers to be more explicit about their interpretive inferences, demystifying the research process and contributing to core goals of transparency and replicability. Positivist researchers often downplay or omit the role of interpretation in the research process. This leaves an incomplete record of how scientific advances are achieved. Discussions of the moment(s) of interpretive insight that sparked a theory or refined an existing one could offer insight into theory generation—an underexamined part of the research process. This advice aligns with calls in other disciplines for more accurate reporting of "revelatory moments" during fieldwork. Trigger et al. (2012, 516-525) define these moments as unplanned and "intense subjective experiences" that generate theoretical insight, and argue that descriptions of these experiences can demystify the process by which researchers arrive at their understandings.

Researchers stand to benefit in peripheral ways from an expanded methodological repertoire that includes traditionally interpretivist methods. We find that our work is more creative with these methods at our disposal. Timothy Pachirat writes that interpretivism is "especially hospitable to the work of imaginative theorizing, of crafting genuinely new and exciting ideas, of nourishing the "playfulness of mind" so necessary to the goodness of social science" (Pachirat 2006). We believe that an interpretive "playfulness of mind" encourages the discovery of the kind of problems and insights that make for creative positivist research. By letting interlocutors' meanings guide parts of our inquiry, researchers can uncover new and puzzling patterns of variation, develop novel insights and explanations, devise creative research designs, and draw insightful and accurate conclusions from their analyses. Pachirat writes that interpretive approaches can help open "that black box that every researcher at one time or another must confront: How do I get/find/have a good idea?" Creativity often arises from new combinations of things; positivists considering how to incorporate interpretive research practices have ample opportunity for new methodological combinations. Some of these may be dead-ends, but some, like Wood's map-making exercise, will introduce new methodological techniques to the discipline.

The "playfulness of mind" that leads to creative research can also produce more memorable and impactful writing.[5] While researchers writing in a positivist framework may be hesitant to genre-bend like some interpretivists,[6] an interpretive orientation can encourage beneficial departures from disciplinary norms. Indeed, the types of data from which interpretive insights are drawn—ethnographic field-notes, stories, poems, jokes, visual and speech data, and so on—are conducive to rich description, and to writing that communicates the distinct character, or "feel," of the context of research.[7]

---

[5] In our experience, colleagues rarely forget reading Pachirat's (2012) vivid description of death and its obscuration in an industrial slaughterhouse. For Pachirat's discussion of his choice of writing style, see *Every Twelve Seconds* (2012, 18-19).

[6] Pachirat (2017), for instance, structures his methodology textbook as a stage-play script.

[7] Nielsen (2017, 1, 27, 169), for instance, attempts to bring interpretations from ethnographic material to the fore in several chapters before proceeding with the staid prose common in positivist research reports.

Importantly, decisions about genre and form are not just presentational: deductive and linear academic writing can downplay the complexities of the social world. Writing that captures the contradictions, ambiguities, and multiplicities of meaning might more faithfully represent the social world it aims to describe. By giving "unapologetic priority to the meaning making of its subjects," an interpretive orientation is "uniquely situated to strengthen the voices and visibility of those who often go unheard and unseen" (Pachirat 2006, 377). Writing that is guided by this orientation might grant more legitimacy to the experiences and understandings of the researched.[8]

Yet as we advocate for researchers to use methods derived from both positivist and interpretivist traditions to address the problem of interpretive inference, we anticipate several concerns. The most commonly stated concerns are that interpretive and positive methods are incompatible, either philosophically or practically. These critics argue that interpretivist and positivist methods can't be meaningfully mixed because of these philosophical contradictions (see, for instance, Schatz 2009, 18, and Ahmed and Sil 2012). Schwartz-Shea and Yanow (2012), for instance, write that, "[i]t is hard, if not impossible, to square [interpretivist] research that rests on constructivist ontological presuppositions and interpretive epistemological ones with [positivist] research that rests on realist ontological and objectivist epistemological ones." A full defense of why these approaches can be compatible is too long for this paper, and we have taken it up elsewhere (English and Nielsen 2024). But we believe that a realist view of science, particularly as articulated by Godfrey-Smith (DATE) can render interpretation and positivism compatible.

As for the question of practicality, we believe that once scholars better understand the benefits, they will be more motivated to learn interpretive methods and that these methods are not out of reach. As with any technique, researchers should learn about methods that are new to them and practice using them. For quantitative techniques, such as text analysis, methodology texts, instruction, and computer code provide a starting point (Grimmer, Roberts, and Stewart

[8] For examples of interpretive research that does exactly this, see, e.g., Katherine Cramer's elevation of rural perspectives in *The Politics of Resentment* (Cramer 2016), and Samantha Majic's attention to the political struggles of sex workers in *Sex Work Politics* (Majic 2014).

2022). In contrast, training in interpretive political science typically happens through practice, rather than instruction. There are some textbooks, but interpretive methods tend to be "taught and learned inductively" (Yanow and Schwartz-Shea 2006, xiii).[9] Even some key interpretive methodology texts are not explicit. Pachirat, for instance, writes *Among Wolves* as a play with a lively back and forth over key debates, but few definitive statements of what good research is or is not. The standards of interpretive research are often left implicit, and, as a result, it takes time to learn interpretive evaluative criteria. The steps required to meet these criteria, moreover, are not always clear for positivists trained on statistics textbooks. Because training in the interpretive tradition occurs primarily through emulation and practice, the way out of this challenge is through it. Positivists who wish to incorporate interpretative methods in their work should do so and then solicit criticism from experts to correct novice mistakes.

Interpretivists rankle at the idea that interpretive methods are only useful if they contribute to causal inference. Hopf, for example, is annoyed at the idea that ethnography should serve as the "summer intern" to the "senior partners" of formal and statistical analysis (Hopf 2006, 18). Pachirat's ethnography textbook criticizes the view that while "ethnography is immensely useful for generating hypotheses, exploring peculiar residuals that appear in statistical analyses, or helping the researcher uncover potential causal mechanisms linking dependent and independent variables," it "must be subsumed within a broader research program in which the other two legs of the stool—statistical and formal analysis— serve to test, and ultimately verify or falsify, the hypotheses and hunches developed by fieldwork" (Pachirat 2017, 16). We are not trying to make interpretivism be the "summer intern" for causal inference. Just because interpretive methods can contribute to causal inference, doesn't mean they must. Scholars focused on causal inference would make better inferences if they incorporated interpretive methods when making interpretive inference, perhaps as a "summer intern" or perhaps as something more. But we aren't asking interpretivists to do anything different, and certainly not to work as interns for positivists. Ethnography and other interpretive methods are

[9] Yanow and Schwartz-Shea, for instance, write that "[e]thnographic and participant-observer research methods in particular have largely been learned through a kind of apprenticeship, through reading others' work in a series of courses and a kind of trial-and-error learning by doing (the "drop the graduate student in the field and see if he swims" sort of teaching)" (Yanow and Schwartz-Shea 2006, xiv).

valuable without being subsumed into formal models for causal inference. Our advocacy for interpretive methods comes from respect for what they can uniquely contribute to causal inference, not from a desire to discipline or diminish interpretivism into only serving causal inference.

5 Conclusion

In this paper we have argued that interpretive inference faces a fundamental challenge: meaning can only be measured indirectly, through representations that are necessarily incomplete. Because researchers can never be certain they have reconstructed an identical representation of meaning, measurements of meanings are prone to error. Yet error in measurement causes serious problems for causal inference if unaddressed, because it violates the assumptions of the prevailing techniques for solving the fundamental problem of causal inference: that we cannot observe outcomes under multiple treatment statuses for the same unit. Every scholar who learns about the fundamental problem of causal inference should also learn about the fundamental problem of interpretive inference and how these problems combine. Social scientists can't eschew the study of meaning, or the study of causes, so an understanding of both is essential.

We have focused on the challenge of interpretive inference when it presents straightforward challenges to causal inference because interpreted meanings are either a treatment or a moderator of a treatment. Measurement error in other variables can induce problems for commonly-used causal inference methods, however, so these examples are just a subset of situations where the problem of interpretive inference must be surmounted to make credible causal inferences.

Holland's (1986) articulation of the fundamental problem of causal inference was not the final word on the problem, and we do not expect to be the final word on the problem of interpretive inference. Rather, by pointing out the problem and showing that it is both fundamental and consequential, we hope to spur further methodological research that addresses it. We hope that scholars will pick up our ideas about how to make robust interpretive inferences run with them. As we have seen, the problem is an old one and numerous approaches exist already, in both positivist and interpretivist traditions. Without the overarching framework,

however, these approaches are known primarily within siloed research communities. We think research practice – trial and error – is the key to progress on interpretive and positive combinations. In this, we follow a long interpretivist tradition of learning by doing. As scholars who are drawn to both approaches feel permission to pursue combinations, they will likely pioneer creative new approaches that we cannot yet imagine, resulting in unforeseen discoveries, facilitated by the integration of interpretive and causal inference.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen. "Analyzing causal mechanisms in survey experiments." *Political Analysis* 26.4 (2018): 357-378.

Ahmed, Amel, and Rudra Sil. "When multi-method research subverts methodological pluralism—or, why we still need single-method research." *Perspectives on Politics* 10, no. 4 (2012): 935-953.

Allina-Pisano, Jessica. "How to tell an axe murderer: An essay on ethnography, truth, and lies." *Political ethnography: What immersion contributes to the study of power* (2009): 53-73.

Benstead, Lindsay J., Amaney A. Jamal, and Ellen Lust. "Is it gender, religiosity or both? A role congruity theory of candidate electability in transitional Tunisia." *Perspectives on Politics* 13, no. 1 (2015): 74-94.

Clarke, Kevin A., and David M. Primo. A model discipline: Political science and the logic of representations. Oxford University Press, 2012.

Corstange, Daniel. "Religion, pluralism, and iconography in the public sphere: Theory and evidence from Lebanon." *World Politics* 64.1 (2012): 116-160.

Cramer, Katherine J. *The politics of resentment: Rural consciousness in Wisconsin and the rise of Scott Walker*. University of Chicago Press, 2016.

Dafoe, Allan, Baobao Zhang, and Devin Caughey. "Information equivalence in survey experiments." *Political Analysis* 26, no. 4 (2018): 399-416.

DeSante, Christopher D. "Working twice as hard to get half as far: Race, work ethic, and America's deserving poor." *American Journal of Political Science* 57, no. 2 (2013): 342-356.

Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. "How to make causal inferences using texts." *Science Advances* 8, no. 42 (2022): eabg2652.

Elder, Elizabeth Mitchell, and Matthew Hayes. "Signaling Race, Ethnicity, and Gender with Names: Challenges and Recommendations." *The Journal of Politics* 85, no. 2 (2023): 764-770.

Fong, Christian, and Justin Grimmer. "Causal inference with latent treatments." *American Journal of Political Science* 67, no. 2 (2023): 374-389.

Holland, Paul W. "Statistics and causal inference." *Journal of the American statistical Association* 81, no. 396 (1986): 945-960.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 905-949.

Lake, David A. "Theory is dead, long live theory: The end of the Great Debates and the rise of eclecticism in International Relations." *European Journal of International Relations* 19, no. 3 (2013): 567-587.

Landgrave, Michelangelo, and Nicholas Weller. "Do name-based treatments violate information equivalence? Evidence from a correspondence audit experiment." *Political Analysis* 30, no. 1 (2022): 142-148.

Pachirat, Timothy. *Among wolves: Ethnography and the immersive study of power*. Routledge, 2017.

Paluck, Elizabeth. "The promising integration of qualitative methods and field experiments." *The ANNALS of the American Academy of Political and Social Science* 628, no. 1 (2010): 59-71.

Pryzant, Reid, Dallas Card, Dan Jurafsky, Victor Veitch, and Dhanya Sridhar. "Causal Effects of Linguistic Properties." In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4095-4109. 2021.

Rasmussen, Stig Hebbelstrup Rye, Alexander Bor, Mathias Osmundsen, and Michael Bang Petersen. "'Super-unsupervised'classification for labelling text: online political hostility as an illustration." *British Journal of Political Science* 54, no. 1 (2024): 179-200.

Rathbun, Brian C., Christopher Sebastian Parker, and Caleb Pomeroy. "Separate but Unequal: Ethnocentrism and Racialization Explain the "Democratic" Peace in Public Opinion." *American Political Science Review* (2024): 1-16.

Ridge, Hannah M. *Defining democracy: Democratic commitment in the Arab world*. Lynne Rienner Publishers, 2023.

Samii, Cyrus. "Causal empiricism in quantitative research." *The Journal of Politics* 78, no. 3 (2016): 941-955.

Sen, Maya, and Omar Wasow. "Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics." *Annual Review of Political Science* 19 (2016): 499-522.

Simmons, Erica S., and Nicholas Rush Smith. "Comparison with an ethnographic sensibility." *PS: Political Science & Politics* 50, no. 1 (2017): 126-130.

Simmons, Erica S., and Nicholas Rush Smith, eds. *Rethinking comparison*. Cambridge University Press, 2021.

Schaffer, Frederic Charles. "Ordinary language interviewing." *Interpretation and method: Empirical research methods and the interpretive turn* (2006): 150-160.

Schwartz-Shea, Peregrine, and Dvora Yanow. 2013. *Interpretive research design: Concepts and processes*. Routledge.

Taylor, Charles. "Interpretation and the sciences of man." Philosophy Education Society, 1971.

Tomz, Michael R., and Jessica LP Weeks. "Public opinion and the democratic peace." *American political science review* 107, no. 4 (2013): 849-865.

Yanow, Dvora, and Peregrine Schwartz-Shea. 2006. *Interpretation and method: Empirical research methods and the interpretive turn.* Routledge.