

## ANALYTICAL ESSAY

---

# The Great Divides: Incommensurability, the Impossibility of Mixed-Methodology, and What to Do about It

DEREK BEACH AND JONAS GEJL KAAS   
Aarhus University

There is still significant confusion about how multimethod research can be undertaken and even if it is possible. The article makes the claims that much of the confusion is the result of a failure to distinguish between *multimethod* and *multimethodology* research. We argue that there are at least three different methodological languages: variance-based, case-based, and interpretivist. The article starts by discussing the ontological and epistemological foundations underlying the three different methodologies that result in them making very different types of claims evidenced with very different empirical material. Variance-based methodologies assesses mean causal effects across a set of cases, whereas case-based methodologies focus on how a causal process works within a case. Markedly different from the causally oriented variance- and case-based approaches, interpretivist research ask questions about human meaning-making in specific contexts. While the claim of methodological incommensurability is not a new claim, the contribution we make in this article is to unpack more clearly the irreconcilable differences that exist across the three methodologies and how they play out in international studies scholarship, and to provide suggestions for what we can do about it.

**Keywords:** case-based, causal mechanisms, counterfactuals, interpretivism, meaning-making, methodology, multimethod research, variance-based

---

Multimethod research has gained in popularity in recent years for studying complex subjects like international studies (IS) (Lieberman 2005; Schneider and Rohlfing 2013, 2016; Humphreys and Jacobs 2015; Seawright 2016; Goertz 2017; Beach and Rohlfing 2018).<sup>1</sup> The promise of multimethod research is that different methods can compensate for each other's relative weaknesses, enabling more robust inferences to be made about the world. There is even a widespread perception among many scholars that high-quality book-length work *has* to engage in multimethod research on a topic.

Yet there is still considerable confusion about how multimethod research can be undertaken and even if it is possible. In this article, we contend that much of the

---

<sup>1</sup> The term mixed-method is also frequently used to refer to essentially the same thing.

confusion is the result of a failure to distinguish between *multimethod* and *multimethodology* research. Research *methods* are particular tools and techniques for analyzing the social world, whereas a research *methodology* defines a family of methods that share similar foundational ontological and epistemological assumptions.

Drawing on recent developments in the philosophy of science and within social science methodology, we claim that there are three fundamentally different methodologies within the social sciences: (1) variance-based methodology, where counterfactual causal claims act as the shared ontological bases that are studied with cross-case evidence of difference-making; (2) case-based methodology, which focuses on causal mechanisms or processes as the fundamental ontological claims that are evidenced using observational traces (aka mechanistic evidence), often supplemented with bounded cross-case comparisons; and (3) interpretivist methodology, which is focused on meaning-making within particular social contexts. Case-based and interpretivist methodologies are sometimes termed “qualitative,” but this is an unhelpful term that tends to conflate two approaches with fundamentally different ontological and epistemological foundations (Yanow and Schwartz-Shea 2012, 2014).

Multimethod research involves combining two or more methods from *within* a single methodological family, for example by coupling an experiment that assesses the causal effect of X on Y within a controlled setting together with a statistical analysis of correlations of X and Y from surveys of a larger sample of the population. In contrast, multimethodological research would involve combining methods from *different* methodologies, for example combining a variance-based method like a controlled experiment with an interpretive analysis of how individuals in a particular context make sense of a given phenomenon.

In this article, we put forward the claim that while multimethod research is possible, multimethodology research is a *fata morgana* because the different methodologies ask fundamentally different questions that are evidenced using very different types of empirical material. For example, if we are studying the phenomenon of civil war, a variance-based design could investigate the covariation of potential causes like economic inequalities between ethnic groups with the outbreak of intrastate war across a large number of countries for individual years. A case-based method like process tracing might select a particular case of an outbreak of intrastate war and investigate the mobilization process of an ethnic group in the conflict, whereas an interpretivist ethnographic study might inquire how villagers who had lived their whole lives together with neighbors of a different ethnic group made sense of the use of violence against them. As can be seen from this example, we would be asking fundamentally different questions—average covariational effects across many cases, mobilization processes within a single case, meaning-making in a particular social context—that are evidenced using very different empirical methods.

The only way that multimethodology research is possible is if a method from one of the methodologies is transformed into a pale shadow of itself by taking less seriously their fundamental ontological assumptions and the epistemological implications of these assumptions, resulting in a form of *methodological monism* in which methods from other methodologies are shoehorned to fit within one chosen methodology. This type of methodological monism is most famously seen in King, Keohane, and Verba’s *Designing Social Inquiry* (1994), in which they claim there is only one logic of social inquiry. They then contend that an ontological understanding of causation as a combination of regularity and counterfactuals should act as the foundations for case study methods, which can be undertaken by transforming case studies into quasi-experimental assessments of difference-making *across* cases.<sup>2</sup> Monism is also found among proponents of other methodologies—for example,

<sup>2</sup> For similar arguments, see Gerring 2011. For critical discussions from a case-based side, see Ragin 2000; Brady and Collier 2011; Goertz and Mahoney 2012; Beach and Pedersen 2016.

witnessed in Bevir and Blakely's claim in *Interpretive Social Science* (2018, 1) that, "anyone who really desires to understand and explain human agency must adopt a 'hermeneutic' or interpretive philosophical perspective," in effect arguing for shoehorning methods from other methodological approaches into an interpretive worldview.

In monist multimethod research, methods from other methodologies therefore become adjunct tools used to illustrate claims that are more strongly evidenced using methods drawn from the methodological mothership. The most common form of this in practice is when a variance-based regression analysis of observational, large-N cross-case data is followed by a handful of illustrative studies of particular cases that serve as *descriptive* exemplars of the cross-case trend.

In this article, we suggest that IR scholars should stop trying to engage in mission impossible but instead learn to take fundamental ontological and epistemological differences more seriously across different methodologies. The three methodological approaches diverge on a set of core ontological and epistemological assumptions that result in them producing very different types of evidence of different types of claims, thereby making it impossible to claim that they can be combined in a form of methodological triangulation to compensate seamlessly for each other's relative weaknesses.

The claim of methodological incommensurability is not a new claim (e.g., [Ahmed and Sil 2009](#); [Chatterjee 2011](#)). However, the contribution we make in this article is to unpack more clearly the irreconcilable differences that exist across the three methodologies based on recent advances within the three approaches and to provide suggestions for what we can do about it when studying IS.

This article starts by introducing the fundamental ontological and epistemological assumptions of different methodologies by distinguishing between variance-based, case-based, and interpretivist approaches. For each, we evaluate the epistemological consequences that differing fundamental ontological assumptions have and provide examples of IS scholarship within the approach. This is followed by a section where we discuss why these differing foundations make the three approaches methodologically incommensurable. The article concludes with a discussion of the challenges that these differences create for social science research, arguing for an appreciation that there is not one "gold standard" method but, instead, that *each of the methodologies* has its *own parallel evidential hierarchy*. We conclude with the recommendation that all social science scholars should learn about the basic foundations of all three methodologies in order for them to have a minimum level of methodological fluency that would enable them to understand what findings mean across methodologies.

### Variance-Based Methodology—Controlled Comparisons and Mean Causal Effects

Variance-based methodology can be understood as a "top-down" type of research, where the focus is on assessing mean causal effects across a set of cases. In variance-based approaches, the evidential gold standard is a randomized controlled experiment, which if properly designed to ensure strong control of potential confounders, enables strong *causal* inferences to be made about the mean causal effect of a given treatment variable *across* cases within the studied sample because of the manipulation of the presence/absence of the cause and the absence of other potential causes (aka confounders) in the controlled setting ([Gerring 2011](#); [Clarke et al. 2014](#); for a more critical interpretation, see [Deaton and Cartwright 2018](#)).

Strictly speaking, the level at which causes are operative is always *within* a single case, with cases understood as an instance of a causal relationship playing out. For example, a sharp decrease in the popularity of a particular US president can theoretically result in her choice to engage in a minor diversionary war in order

to bolster support due to a potential “rally round the flag” effect (e.g., [Brulé 2008](#)). However, the decline in popularity of a given president will not impact the war-making choices of other presidents (i.e., in other cases), unless there are important learning effects in subsequent cases over time. If learning effects do work across cases, the individual cases are no longer independent but, instead, are episodes in a sequence that, as a whole, is one case of learning from past behavior. If learning does not bind cases, then the cause (decline in public support) is potentially linked to the outcome (choice to engage in a diversionary war) *within* a given case, not *across* cases. Therefore, a very important distinction can be made in variance-based research: while causation *always* takes place *within* cases, we *learn* about causation by studying variation *across* cases. For example, this could be done by assessing the effect on values of the outcome (aka dependent variable) of variation in the level of the cause (aka independent variable), controlling for other potential causes of variation across a number of cases.

The variance-based approach is built on a *counterfactual* understanding of causation—often termed the potential outcomes framework—which is situated firmly within a neopositivist understanding of science ([Woodward 2003](#); [Rubin 2005](#); [Angrist and Pischke 2009](#); [Jackson 2016](#)). Counterfactual causation is a claim that a cause can produce an outcome when we are able to substantiate that the cause’s absence would result in the absence of the outcome, all other things held equal (e.g., [Lewis 1986](#), 160; [Woodward 2003](#)). When working with counterfactual claims, without evaluating the *difference* that a cause can make in a case where it is present and the counterfactual absence case in a controlled setting, no *causal* inferences are possible.

The need for controlled comparison and the difficulties achieving it relate to what Holland termed the *fundamental problem of causal inference* ([Holland 1986](#)), in which we ideally would assess the effect of the treatment and absence of it on the *same* case by rolling back time. For example, we would ideally compare a case in which an unpopular president engaged in a minor diversionary war with an alternative reality in which the only difference was that the president maintained her popularity to assess whether the president in the counterfactual case also engaged in the minor war. As this type of counterfactual comparison of potential outcomes is impossible in the complex real world, we are forced to engage in a second-best assessment of the counterfactual causal claim by comparing the impact of the treatment and its absence *across a set of cases* (aka potential outcomes), holding the impact of all other potential causes of difference as constant as possible. The need to compare across cases to assess the difference-making of counterfactuals means that the potential outcomes framework can also be termed a form of “top-down” analysis, in contrast to a “bottom-up” assessment that explores causal relationships within individual cases ([Illari and Williamson 2011](#)).

Our ability to compare potential outcomes in treated and nontreated (counterfactual) cases in a controlled setting is greatest in an experimental design. In a randomized controlled experiment, we compare the values of a group of cases that receive a treatment with another group that is untreated in which there are (hopefully) no other differences between the groups than the treatment. If we then can measure substantial differences between the values of the outcome in the treated and nontreated groups, this is the mean causal effect of the treatment.

To be able to use controlled comparisons to make causal inferences, two assumptions have to hold in our design: unit homogeneity and independence of units ([Holland 1986](#); [King, Keohane, and Verba 1994](#), 91–97). *Unit homogeneity* is when a cause produces the same outcomes in two or more cases. *Independence of units* can be defined as the situation in which the potential outcome in one case is unaffected by values of the cause in other cases. If these two assumptions do not hold in a study, biased estimates of the difference that variations in X have for values of Y are produced. For instance, if the units are not homogenous,

there might be important differences between cases that can act as *confounders* (aka alternative causes), whose presence prevents us from determining whether the cause or confounders produced differences in values of the outcome across cases.

In variance-based research, the degree to which these two assumptions hold in different designs results in a clear *evidential hierarchy* that relates to the strength of *causal* inferences that can be made within the given study (aka *internal validity*) (Gerring 2011; Clarke et al. 2014).<sup>3</sup> Experimental designs are at the top, although there is a clear distinction between laboratory and real-world experiments. Controlled laboratory experimental designs are at the top of the evidential hierarchy because the controlled laboratory setting means that there is a lower risk of unknown confounders that could produce differences in comparison to both survey and field experiments. While there might be differences in how causes work between individuals in the study, the random assignment of a large number of cases into treatment and control groups ensures that the units in our study *are on average* homogeneous, with individual differences washing out. Unit independence is also best ensured in a controlled experiment, where random selection ensures that the values of X are independent of values taken by Y. As an example, Tomz and Weeks (2013) undertook a survey experiment testing the democratic peace thesis using a public opinion poll deployed in the United States and UK, in which they asked a large number of respondents whether they would support a preemptive strike on a country that could develop nuclear weapons within six months. The experimental manipulation related to giving randomly selected respondents different information about the country's political regime (democracy or autocracy) and controls such as military alliances and power. They found that—other things equal—respondents who were told that the target country was democratic were on average less likely to support military intervention. They conclude that “our experimental approach allows us to conclude with confidence that the effect of democracy is genuinely causal” (Tomz and Weeks 2013, 862).

When we move away from experimentally manipulated designs to purely observational data, we move down the evidential hierarchy relating to the internal validity of our causal claims. Many scholars claim that causal claims are still possible when the observational data gained from a large number of cases has the character of a *natural experiment*, where we are able to substantiate that the population is split (either temporally or spatially) into a treatment and control group in which everything else is constant—meaning that we can assume that the cases were assigned into the treated and control groups randomly by nature (Angrist and Pischke 2009; Gerring 2011; Dunning 2012).

When we move beyond natural experiments to other forms of large-N comparisons using observational data, it becomes increasingly difficult to make *causal* inferences because of the increased risk of confounders that cannot be controlled for. Some IS scholars suggest though that we can make cautious causal inferences on topics like the democratic peace with controlled comparisons using large-N data sets. In the words of Dafoe, Oneal, and Russett (2013, 202), “the study of international relations can rarely generate strong causal inferences about important questions . . . However, the analysis of cross-country data, done with care, can provide descriptive information about the size of conditional associations and how much they stand out from the unexplained variance (their statistical significance). Cautious causal inferences can then be drawn.” Here, it is the ability to control for potential confounders across a large number of cases that enables cautious causal

<sup>3</sup> In medicine there exist many explicit attempts to produce evidential hierarchies (e.g., the Oxford Centre for Evidence Based Medicine, <https://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>). Interestingly, at the top of the evidential hierarchy in medicine are metastudies of multiple randomized controlled experimental studies. Obviously, if findings are robust across different design specifications and sample populations, then the internal and external validity of the evidence is much stronger than from a single study.



inferences to be made, although other scholars would contend that the weaker ability to control for confounders makes the causal claims shaky at best.

Further down the evidential hierarchy are small-N comparisons, where the small number of cases makes it increasingly difficult to maintain that the assumptions of unit homogeneity and independence hold, meaning that there can be many different confounders that are not controlled for. Unit heterogeneity can be the result of differences between cases, which is an endemic problem in IS for small-N comparisons because there will always be a large number of potential differences across a small number of all but the most simple types of cases (Liebersohn 1991). Ideally, we would be able to identify two cases that are similar in all respects except for the independent variable of interest, but because of the complexity of the IS phenomena that we study, it is difficult to find cases in which the “all other things equal” assumption required for causal inferences actually holds (Ragin 1987, 48). One prominent scholar, who otherwise advocates assessing counterfactuals in small-N designs, admits that, “Controlled comparison and matching face the difficulty of finding real-world cases that are identical in all respects but one” (Levy 2015, 390). This means that there will realistically always be multiple differences between a treated and nontreated case that could have produced different outcomes. As a result, small-N comparisons are far from the experimental ideal because the lack of control means that we have very weak evidence of counterfactual causation, at best.

At the very bottom of the evidential hierarchy in variance-based methodologies are single-case studies. Given the need to assess the counterfactual, single cases in a variance-based potential outcome framework need to be transformed from a single case into a multicase comparisons. One strategy for achieving this is proposed by King, Keohane, and Verba (1994, 217–28), where they famously argue for transforming “one case into many” by disaggregating a case either spatially or temporally, in effect creating a small-N comparison out of a single case. Unfortunately, transforming “one case into many” results in a design where the units are neither homogeneous nor independent. One way to transform one case into many involves observing variation in the dependent and independent variables over time, thereby creating multiple units to assess difference-making. This could involve disaggregating an international negotiation such as COP-15 into multiple units over time ( $t_0, t_1, \dots, t_n$ ). However, as they readily admit, the resulting units would not be independent of each other because what happens at the start of a negotiation would naturally impact events later in the negotiation, meaning that “When dealing with partially dependent observations, we should be careful not to overstate the certainty of the conclusions” (King, Keohane, and Verba 1994, 222). Additionally, the units over time might not be causally homogeneous because a given cause such as discursive power might work differently during an agenda-setting phase dominated by deliberative dynamics and an endgame dominated by power politics.

Another way of doing variance-based single-case studies involves using counterfactual single-case studies, where a comparison is undertaken between the actual case and an alternative “what might have been” case in an alternative hypothetical universe in which the cause is not present (Fearon 1991; Tetlock and Belkin 1996; Lebow 2000; Goertz and Levy 2007; Levy 2015). However, because the counterfactual comparison case is purely hypothetical, there is no actual evidence that can enable us to assess whether things really would have been different if the cause was not present, meaning that while a hypothetical comparison can be useful as a thought experiment, it is at the bottom of an evidential hierarchy in variance-based research.

Taken together, this means that in variance-based approaches, case studies should be seen as adjunct tools that can be used to illustrate causal effects found in large-N designs; to detect potential causal mechanisms that can then be tested empirically using more robust, large-N cross-case analysis using mediation analysis; to explore whether a natural experiment is actually a natural experiment; or to detect

measurement error or potential confounders. In contrast, the core causal inferences are made using large-N controlled comparisons—ideally using experimental designs.

In conclusion, variance-based approaches can be viewed as top-down methods for the assessment of counterfactual causation using controlled comparisons across cases. Relative strengths include the ability to assess the magnitude of net causal effects across cases and the ability to make causal inferences about many cases (populations or samples thereof), especially when we are able to use experiments. However, the downside is that the more a study approximates a controlled experiment, the less it typically approximates real-world settings and the lower our ability to generalize beyond the studied set of cases. In variance-based research, there is therefore usually a trade-off between the *internal* validity and *external* (i.e., ability to infer to other cases outside of the studied cases) and *ecological* (i.e., whether the findings gained from the research setting hold outside of the design) validity. Lab or survey experiments have a high level of internal validity, but they typically have low external and ecological validity. For example, a survey experiment of the democratic peace (e.g., [Tomz and Weeks 2013](#)) might have high internal validity, but the results cannot necessarily be generalized to other settings (e.g., the United States in the 1960s or other democratic countries that have very different political systems, such as Japan) (i.e., low external validity). There can also be problems related to whether the treatment given in an experiment is analogous to real-world settings. In relation to Tomz and Weeks' study, it is questionable whether the results gained from asking respondents in a survey about a hypothetical case actually are transferable to a real-world crisis situation in which we should expect a range of other factors to be at play (e.g., media and partisan priming). Natural experiments have lower internal validity because we are *assuming* that the situation approximates an experiment, but there is typically higher ecological validity because of the real-world setting. Outside of experimental or quasi-experimental designs, the internal validity of causal claims is quite low. In contrast, large-N comparisons using observational data typically have higher external and ecological validity, whereas small-N comparisons and single-case studies have neither internal nor external validity.

### Case-Based Methodology—How Causes Work in Cases

The core of case-based methodology is *within-case* tracing of causal processes (also termed mechanisms)<sup>4</sup> using detailed case studies, although cross-case comparisons are important for selecting appropriate cases and helping to generalize findings about processes found within particular cases (see [Ragin 2000](#); [Schneider and Rohlfing 2013, 2016](#)). Cross-case comparisons can also be done at the *level of causal processes* by investigating whether particular parts of a process (e.g., a critical juncture in which a leader makes a decision to mobilize for war) are present in two or more cases and then exploring the similarities and differences in how the subsequent process played out in the different cases (see e.g., [Kruezer 2020](#); [Thelen and Mahoney 2015](#)).

Case-based methodology can be viewed as “bottom-up” research designs, as the analytical point of departure is the in-depth tracing of causal processes in individual cases and how they play out ([Russo and Williamson 2011](#)). Building on realist philosophical foundations (e.g., [Bhaskar 1978](#); [Sayer 2000](#); [Jackson 2016](#)), as the goal is to get as close as possible to what is happening within cases, this means that case-based research is typically very sensitive to the *context* in which causal processes operate ([Ragin 2000](#), 53–63; [Falletti and Lynch 2009](#); [Thelen and Mahoney 2015](#), 7–15). This focus on complexity and causal claims within bounded contexts is echoed in broader developments in other fields of science, including in systems

<sup>4</sup> We use the terms causal process and causal mechanisms interchangeably in this section.

biology and personalized medicine (Ahn et al. 2006; Levi-Montalcini and Calissano 2006; Cartwright 2007, 2012; Bechtel and Richardson 2010; Khosrowi 2019).

Causal processes are viewed as something different from causes themselves in case-based research. Instead, a causal process is what links a cause and an outcome. While there is agreement about the goal of tracing processes, there is not agreement within case-based methodology on the ontological nature of causal processes. Some scholars pragmatically focus more on the epistemological question of how we can trace processes empirically, suggesting that scholars should collect diagnostic evidence that sheds light on “processes, sequences, and conjunctures of events within a case,” without defining the exact ontological nature of causal processes (aka causal mechanisms) (e.g., Brady and Collier 2011; Bennett and Checkel 2014). Other scholars draw on developments in the broader philosophy of science by suggesting that causal processes should be unpacked theoretically into parts composed of entities engaging in activities that bind causes and outcomes together in a productive relationship (Machamer 2004; Machamer, Darden, and Craver 2000; Russo and Williamson 2011). Entities are the social actors who engage in activities. What form they take depends on one’s theory, with entities potentially individuals or macro-entities such as a nation-state. Activities are at the heart of the productive understanding because they are what bind parts of a causal process together. Examples of activities can include using military force, voting, or engaging in a normative speech act, depending on one’s theory.

In both understandings of causal processes, the goal is to understand how a process actually works within a case by tracing its operation—either at the aggregate level or as an unpacked process. Exploring how a causal process worked within a case often requires a back-and-forth, abductive research strategy in which one’s initial theoretical ideas about how the process worked are empirically incorrect (Tavory and Timmermans 2014). This should then lead to a phase of theory-building focused on trying to understand how the process played out, which can subsequently be tested again using new mechanistic evidence.

Irrespective of whether mechanisms are unpacked in detail theoretically, most case-based scholars agree that causal processes / causal mechanisms are something different from counterfactual claims (e.g., Bennett and Checkel 2014; Beach and Pedersen 2019). These arguments are most clearly seen within the productive understanding of causal mechanisms, where scholars like Machamer (2004) and Groff (2011) contend that tracing mechanisms involves exploring what actually took place within a case, not what could have happened in other logically possible counterfactual worlds. In the words of Bogen (2005, 415), “How can it make any difference to any of this whether certain things that did not happen would have or might have resulted if other things that did not actually happen had happened?” Similarly, Groff (2011, 309) claims that mechanisms are real processes that involve the exercise of causal powers in the real world, not in logically possible counterfactual worlds. Taken as a whole, this means epistemically that we are shifting the analytical focus from assessing difference-making across cases to focusing on the actual process that links them together in a productive relationship.

At the epistemological level, case-based scholars agree that the type of empirical evidence that enables within-case causal inferences about processes is the traces left by their operation in a case. While terms such as causal process observations are often used (e.g., Brady and Collier 2011), the term *mechanistic evidence* from the philosophy of science is a more precise term—defined as the observable fingerprints left by the operation of the activities associated with parts of mechanisms (Russo and Williamson 2007; Illari 2011; Beach and Pedersen 2019). In this type of evidence there is *no variation across cases*; instead evidence is constituted of the empirical traces left by activities performed by social entities, which then enable us to infer that there is a causal process within a particular case



(Clarke et al. 2014; Fairfield 2019; Beach and Pedersen 2019). Mechanistic evidence is therefore observational data that enables causal inferences to be made *within* cases. When understood in these terms, the fundamental problem of causal inference is therefore *only* fundamental for variance-based approaches. While evidencing a counterfactual claim requires per definition the assessment of difference-making using controlled comparisons *across* cases, process tracing enables strong causal inferences about causal processes to be made by exploring what actually took place *within* a case using observational, within-case mechanistic evidence.

In case-based research, the detailed tracing within individual cases of unpacked causal processes using mechanistic evidence is at the top of the evidential hierarchy (Clarke et al. 2014). As an example, O'Mahoney used an in-depth case study analysis to analyze how states try to win over undecided states in support of their own position through what he terms rhetorical adduction (2017). Using the conflict between India and Pakistan in 1971 and the issue of the recognition of Bangladesh after the Indian invasion, he first unpacks the process theoretically into four parts: (1) opposition argued that policy is X; (2) claimant argues policy is not X, it is Y; (3) action is taken by claimant that links policy with Y; (4) audience either resists, acquiesces to, or supports policy. He then develops a number of observable implications associated with the parts that are assessed in the case. He finds evidence suggesting that the Indian withdrawal from Bangladesh (the action described in part 3) bolstered Indian arguments, producing the outcome of recognition of Bangladesh by the international community. While the author is able to make strong within-case causal inferences about the process, it is outside of the design to be able to infer whether similar processes are present in other cases because they have not been studied.

Below the detailed within-case tracing of unpacked mechanisms are inferentially weaker within-case methods that do not explicitly trace mechanisms (e.g., congruence studies and analytical narratives), which as a result do not enable as strong of causal inferences as detailed process tracing (e.g., Abell 2004; George and Bennett 2005; Beach and Pedersen 2016). For example, Khong (1992, 66–68), in his classic study of analogical reasoning and foreign policy decision-making in the Vietnam War, deployed congruence analysis by assessing whether there was a match between a policymaker's analogy-based beliefs and policy choices at critical junctures. However, he clearly admits that it does not enable causal inferences; it only acts as a double-check on the results of his process tracing (Khong 1992, 66).

At the bottom of the hierarchy are comparisons *across* a limited number of cases using methods like qualitative comparative analysis (QCA) that can be used to find potential causes, select appropriate cases for within-case analysis, and enable cautious generalizations about processes to small, bounded sets of cases but that do not enable causal inferences about process to be made because no within-case processes are traced. Instead, case-based comparisons only enable claims to be made about empirical associations existing between causes and outcomes (for more, see Beach and Pedersen 2016).

Concluding, the relative strength of case-based research is the ability to make inferences about how things work at the level of process *within* particular real-world cases. As with variance-based designs, there is a trade-off between internal and external validity.<sup>5</sup> In-depth process tracing case studies enable strong causal inferences about process to be made within a given case but leave us in the dark as regards whether similar processes are present in other cases. Multiple additional studies in other cases would then be necessary to ensure the external validity of a causal claim that similar processes are at work within the set of studied cases. While the low external validity might seem to be a critical weakness, the contextual sensitivity of

<sup>5</sup> Ecological validity is not a concern in case-based research because real-world cases are investigated.

claims in case-based research is seen by proponents of case-based designs as a critical strength because we learn how things work in particular contexts (e.g., [Ragin 2000](#); [Cartwright 2011, 2012](#); [Goertz and Mahoney 2012](#); [Beach and Pedersen 2016, 2019](#)). In case-based research, comparisons across cases do not enable causal inferences to be made unless they also embody some form of within-case, process-level analysis. When many cases are compared at the level of process, the strength of evidence for a process in any given case is typically quite weak (low internal validity), but the external validity is higher.

### Interpretivist Methodology—Meaning-Making in the Social World

Research aiming at disclosing and elucidating how agents make meaning of the socially complex world can be defined as *interpretivist*. The fundamental move made by interpretivists is to distinguish between the natural and social worlds. Moving away from the causal-oriented natural world, interpretivists claim that the most interesting questions in the social world relate to understanding human meaning-making “by learning what it does and how people use it” in specific social contexts ([Yanow and Schwartz-Shea 2012, 23](#)). Whereas interpretivist researchers debate whether interpretivism reflects an explicitly antinaturalist stance ([Bevir and Blakely 2018](#)), we see a recurring key aspiration among interpretivist researchers to elucidate context-specific meaning that “potentially reveals (or raises questions about) assumed, unspoken or taken-for-granted ideas” rather “than seeking generalized meaning abstracted from particular contexts” ([Yanow and Schwartz-Shea 2012, 23](#)). We return to this aspiration to disclose the “taken for granted” below.

Interpretivism is based on a double-hermetical practice of humans studying humans ([Jackson 2014, 2016](#)). The intellectual roots of interpretivism are firmly rooted in hermeneutics and phenomenology. The works of scholars such as Mark Bevir, Roderick A. W. Rhodes, Dvora Yanow, and Peregrine Schwartz-Shea have propelled the growth of an interpretivist research community (e.g., [Bevir and Rhodes 2010, 2016](#); [Yanow and Schwartz-Shea 2012, 2014](#); [Bevir and Blakely 2018](#)).<sup>6</sup>

Unlike variance- and case-based approaches, the logic guiding interpretivist research is neither *top-down* nor *bottom-up*. Instead, interpretivist research departs from a *relational* logic of inquiry. Examples of relational logic of inquiry found in IS (e.g., [Jackson and Nexon 1999](#); [Lynch 1999](#); [Weldes 1999](#); [Krebs and Lobasz 2007](#); [McCourt 2016](#); [Qin 2018](#)) mainly depart from [Andrew Abbott’s \(1995\)](#) and [Mustafa Emirbayer’s \(1997\)](#) integrative sociological frameworks, as well as [Nicolas Onuf’s \(1989\)](#) rules-rule reconstruction of world politics. These relational approaches differ from prominent IS constructivist [Alexander Wendt’s \(1999\)](#) systemic theory of three cultures of anarchy (Hobbesian, Lockean, and Kantian). Whereas Wendt argues that each culture of anarchy manifests a structural script for states’ interests and actions, [Onuf \(1989, 22\)](#) views rule—*not* anarchy—as the “general property of political society.” Rule is a result of rules constructed, proliferated, and maintained by human agents finding these meaningful ([Onuf 1989, 22](#)).

IR scholars interested in human meaning-making often have constructivism as their theoretical point of departure. However, not all IS constructivists are interpretivists ([Fearon and Wendt 2005](#)). As [Jackson \(2016, 224\)](#) notes, constructivism is “first and foremost a social theory [and] *not* a methodological claim.” In short, within IS there are theoretical constructivists and those who hold constructivist ontological views ([Hopf 1998](#); [Jung 2019](#)). As an illustration of this difference, whereas [Ted Hopf \(2002\)](#) tests if and how certain social identities affected Soviet and Russian foreign policy interests and choices, [Jackson \(2006\)](#) uses a form of textual

<sup>6</sup> Contributions by Mark Bevir, Roderick A. W. Rhodes, Dvora Yanow, and Peregrine Schwartz-Shea build on top of pioneering social science interpretivist work such as *The Social Construction of Reality* ([Berger and Luckmann 1966](#)), “Interpretation and the Sciences of Man” ([Taylor 1971](#)), and *The Interpretation of Cultures* ([Geertz 1973](#)).

ethnography to elucidate the contestations and commonplaces of meanings concerning the post-WWII reconstruction of “Western Civilization” and implications for transatlantic ties in-between German politicians and scholars. Theoretically, Hopf (2002) and Jackson (2006) both draw on social constructivism within IS, but, methodologically, Jackson and Hopf employ very different logics of inquiry, with Hopf’s form of analysis more in line with case-based approaches.

Unlike a substantial-essentialist study focusing on personal motives as beliefs and *mental causes* to action, a relational study focuses on action as influenced by intentions—following Gertrude E. M. Anscombe’s *Intention* (1957)—understood as envisioned outcomes. These envisioned outcomes are not defined in a vacuum but are mediated by rules (more or less explicitly specifying acceptable action) and roles (assigning rules to different positions) within and sensitive to the given context in which they are formulated. Intentions are ascribing goals to certain actions and thereby ascribing action meaning. How these intentions and goals transcend into action depend on how actors intersubjectively negotiate what constitute meaningful action *in-between* them.

The relational point of departure builds on ontological and epistemological underpinnings that differentiate interpretivist research from variance- and case-based approaches. Ontologically, interpretivists reject an understanding of causality in both modern counterfactual or mechanistic terms that produce explanations in the form of *if, why, and how* something “is as it is” (Bevir and Kedar 2008, 515). However, interesting differences about understandings and the use of causal language exist among interpretivists. Whereas some interpretivist—for instance Lene Hansen in *Security as Practice* (2006, 5)—have abandoned causal speak completely in favor of “non-causal” explanations, others have attempted to work with a broader language of causality. Developing a “processual sense” of causality, Lee Ann Fujii, for example, interpreted a relational social world via a nonmechanistic set of “how” questions. For instance, how did “ordinary people come to be involved in mass violence and how did different actions (both violence and non-violent) become possible in different contexts?” (Fujii 2008, 572).

Answering such questions does not entail identifying the mechanism explaining the processes leading to involvement in mass violence across or in a few cases of mass violence but disclosing how these specific “people make sense of the violence” in the specific context of interest (Fujii 2008, 572). To an interpretivist, causal claims are constitutive. Constitutive means that claims are being made about agents and their ways of naming and framing certain phenomena in a specific context and how such meaning-making structures what they say and do in that context in forms that are meaningful to them. The relational meaning-making process among agents’ naming, framing, and acting in context cannot be “disentangled as to render it in stimulus-response form” (Schwartz-Shea 2014, 146). An interpretivist not only rejects disentangling causes and outcomes but is explicitly sensitive to interdependence—denoted intertextuality—as a trademark of scientifically valid research.

Epistemologically, interpretivist research is designed to *trustworthily* make sense out of how specific agents make sense of their social setting, as accessed by the researcher (Yanow and Schwartz-Shea 2012). Interpretivists are keenly aware of the impact of the researcher themselves on what is being studied, defined as positionality. This refers to the socially and historically constructed context researched agents as well as researchers are situated in. Positionality influences how the researcher makes sense of their own experienced reality and the social context they are researching. In order to trustworthily inquire about how agents make sense of their experienced lifeworld, interpretivist researchers have to constantly reflect on how their positionality—in and outside the researched setting—influences their interpretations of the researched agents. The goal is to make clear how knowledge-claims build on the understandings from encounters with—rather than researcher’s pre-understandings about—the researched agents within a social context.

In *The Subjectivity of the “Democratic” Peace*, Ido Oren (1995) argues that the widely held claims about the democratic peace reflect a general negligence of researchers’ will and capacity to reflect on their positionality and the ideals, values, and power structures influencing their construction and use of social science concepts. Oren (1995, 178) concludes that the claim that democracies are not fighting one another is “better understood as a claim about peace among countries conforming to a subjective ideal that is cast [in] America’s self-image” than as a universal law in world politics.

Instead of generalization, contextualization consequently becomes the most important interpretivist research criteria to assess the quality and relevance of evidence used to compile knowledge-claims (Yanow and Schwartz-Shea 2012, chap. 6). Unlike variance- and case-based approaches use of what Clifford Geertz (1974, 28–30) coined *experience-distant* concepts suited for producing knowledge-claims across a large or small number of cases, the formation and use of *experience-near* concepts are essential for interpretivist research to trustworthily make sense of how agents make sense of their lifeworld as well as trustworthily convey interpretivist knowledge-claims to outside-readers (Schaffer 2016).

Interpretivists are well-aware that their knowledge-claims are neither universally “true” nor—necessarily—generalizable to other agents and contexts. Instead, knowledge-claims mirror encounters between a researcher—holding personal, social, economic values and predispositions—and researched agents situated in context. Importantly, the researcher is the main instrument to generate and analyze data in the interpretive research process (Yanow and Schwartz-Shea 2012, 85; Fuji 2017). Dvora Yanow notes that data in interpretive research is generated not by agents “themselves, or the events and conversations and settings and acts, or even the documents, but rather the researcher’s views of these, as encapsulated in her notes . . . human science data are never really ‘raw’ and ‘unprocessed.’” (Yanow and Schwartz-Shea 2014, xxi). We return to what that demands of researcher’s will and ability to critically reflect on exposure to meaning-making by the agents below.

However, before forming *experience-near* concepts and contextualizing meaning-making to ensure the trustworthiness of knowledge-claims, setting and agents of interest need delimitation. Delimitation is a necessary prerequisite to obtain *access* to the meaning-making of relevant agents (Yanow and Schwartz-Shea 2012, chap. 4).

Trustworthily making-sense of how agents understand the social world they experience requires the researcher to be sufficiently exposed to meaning-making. In obtaining access, several practical considerations are worth considering before starting to conduct an interpretive inquiry. Can physical access to research settings (e.g., archives, prison, slaughterhouse, or embassy) be obtained? Is the researcher sufficiently proficient in the language, dialect, slang, or even jargon used by researched agents in order to access their meaning-making? Sometimes researchers may rely on guides in order to access meaning-making. However, using guides is not unproblematic since the trustworthiness of knowledge-claims relies on the researcher making them (Yanow and Schwartz-Shea 2014, xxi; Pedersen 2019, 133–34, 141).

Returning to researcher’s will and ability to critically reflect on exposure, we find what Samuel E. Morison denotes as *mesure* relevant. Morison (1951) defines the French word *mesure* as a principle of “balance or proportion.” In terms of trustworthiness, *mesure* denotes the individual researcher’s will and ability to trustworthily reconstruct, map, identify, and analyze agents’ relational meaning-making in a specific setting, while remembering embedding this specific setting into its interrelation with the wider societal and political contexts surrounding it. Concretely—Morison exemplifies—*mesure* means that one can, “no longer write political history without considering social forces, or social history without describing political acts and conditions that translate aspirations into deeds” (Morison 1951). Morison concludes there is “no royal road” to achieve a sense of *mesure*. Although “a becoming

humility toward his fellow workers, and skepticism directed toward himself as toward them, will be of assistance.”

Consequently, there is no universally applicable answer to what constitutes the gold standard method to generate and analyze data interpretively after access have been obtained. To an interpretivist, any—or a combination of—methods enabling trustworthy thick descriptions of specific agents in a given setting are welcome. In line with Morison’s recommendation of “a becoming humility,” a good general starting point for producing thick descriptions is *listening* (Pedersen 2019, 123). Listening carefully constitutes in principle an infinite, iterative four-step hermeneutical process of *cultivating*, *encountering*, *writing*, and *presenting* meaning-making of relevant agents. Through this iterative process, the researched meaning-making gradually reveals itself to the researcher on its own terms—however, without the researched ever revealing itself completely to the researcher.

Underlying this iterative four-step hermeneutical process is an *abductive* way of reasoning (Yanow and Schwartz-Shea 2012, 27–34, 2014, xviii–xix). Whereas abduction in a case-based approach involves a form of “inference to the best explanation” (Tavory and Timmermans 2014), in interpretive research it involves a process whereby the researcher become gradually more embedded in the researched setting—the researcher becomes able to better make sense out of the agents’ meaning-making. Unlike variance-based approaches, interpretivist research designs are intentionally flexible to accommodate the understandings about the researched that gradually emerge as the researcher’s preunderstandings are engaged in the iterative process of going through the four-steps consisting of both field- (cultivation and encounter) and deskwork (writing and presenting). From an abductive point of view, the researcher simultaneously “maintaining a balance between being a stranger and being a familiar” to a researched site and agents is central to producing valid interpretivist knowledge (Yanow and Schwartz-Shea 2012, 29). Knowledge is generated in the nexus between “exceptions to” and “experience gained from” encounters between researcher and researched.

Critical when cultivating is the researcher’s “particular competencies and skills to maneuver effectively . . . [and] adapt to the field setting” (Yanow and Schwartz-Shea 2012, 74). The key purpose of cultivation is ensuring that the researcher becomes sufficiently exposed to encounter the complex web of meanings expressed by agents in the setting. Having cultivated the researcher to the researched, research moves on to the encounter between the researcher and researched. Encounter constitutes an iterative process of observing and reflecting. After the encounter comes writing. The core aim of the writing is becoming aware of what the researcher knows and—more importantly—what the researcher still needs to know. Edward H. Carr encapsulated the aim and process of (re)writing, noting, the, “more I write, the more I know what I am looking for, the better I understand the significance and relevance of what I find” (Carr 2001, 23). Having undertaken the initial writing-down of significant events and reactions from relevant agents—based on field notes and initial hunches from cultivation—what Clifford Geertz (1973, chap. 1) coined *thick descriptions* emerge out of the multiple revisions. The thicker descriptions become, the more trustworthy they become as evidence in support of interpretivist knowledge-claims. The final step is presentation. Presenting one’s research orally and/or textually enables public criticism, which is central to demarcating our Weberian understanding of science from nonscience (Jackson 2016, 209). In addition, listening to comments and reviews from peers as well as members of the researched setting<sup>7</sup> sharpens the accuracy and trustworthiness of the presented knowledge-claims.

<sup>7</sup> This is frequently denoted as member-checking (Yanow and Schwartz-Shea 2012, 106–7) or informant feedback (Bygstad and Munkvold 2011).



**Table 1.** Ontological and epistemological differences across the three methodologies

	Variance-based	Case-based	Interpretive
Analytical point of departure	Population-level (top-down)	Case-level (bottom-up)	Individuals (meaning-making in context)
Types of claims	Mean causal effects, building on counterfactual and probabilistic assumptions	How a process works, building on mechanistic and deterministic assumptions. Often combined with claims of necessity and/or sufficiency within bounded comparisons	What does A mean in context B, building on contextual meaning embedded in social relations
Types of evidence	Evidence of difference-making across cases	Mechanistic, within-case evidence and invariant cross-case patterns	The thicker the descriptions, the more trustworthy knowledge-claims
“Gold standard” method	Experiment	Detailed process-tracing + bounded comparisons	Any method that enables trustworthy thick description, e.g., textual and ethnographic methods

**Methodological Incommensurability and What to Do about It**

Table 1 summarizes the key ontological and epistemological differences across the three methodologies in the social sciences.

*Why Are the Methodologies Incommensurable?*

In this section, we discuss why the three methodologies are foundationally incommensurable. We start with discussing the incompatibilities between variance- and case-based approaches because they both operate with explicit causal language and experience-distant concepts. We then discuss why interpretivist research marks an even larger departure from the other approaches. It is important to iterate that we are not claiming that one methodology is better, just that taking their underlying ontological and epistemological assumptions seriously means that they are fundamentally *incommensurable*.

In *variance-based research*, controlled comparisons are used to assess empirically counterfactual ontological claims, ideally through a randomized controlled experiment. Because of the need to compare difference-making across units (ideally a large number), ontologically *probabilistic* claims about *mean* causal effects in a population (or a sample thereof) are made. This means that evidence of mean causal effects is in essence a claim that “it works somewhere” (Nancy Cartwright 2011). However, finding evidence of a mean causal effects is very difficult to communicate meaningfully with findings from single cases, irrespective of whether we are talking about the causal mechanisms operative in a single case or socially embedded meanings within a particular relational context.

If we focus first on the difficulties of communicating between variance-based findings of mean causal effects across cases, and case-based findings of causal processes operative within cases, one core problem is that the assumption of causal heterogeneity within a given population almost never holds (Williams and Dyer 2009, 210–11; Leamer 2010; Cartwright 2011, 2012, 980–81; Khosrowi 2019). If a set of cases were causally homogenous, we could move effortlessly from mean causal effects across cases to individual cases (Rubin 1980, 961; Morgan and Winship 2007,

37–40; Cartwright 2011). However, when this assumption does not hold, averages do not necessarily represent any individual case. Instead, a mean causal effects can be produced by complex interactions between causes and contextual factors—for example, with large positive effects in some cases, smaller positive effects in others, and negative effects in still others (e.g., Leamer 2010). If this were the situation, the set of cases would be far from causally homogeneous, meaning that the mean causal effect is uninformative for any individual case. This is a situation commonly faced even in experiments, where some subjects respond very positively to a treatment stimulus, whereas others do not. As a result, causal effects are only assessed across a relatively large number of units, enabling strong inferences about mean causal effects, but which are not necessarily predictive of any individual in the study.

One commonly suggested way to move from trends across cases to individual cases can be to view causes as “probability-raisers” (Gerring 2011, 199). If our cross-case analysis found a mean causal effect in which higher values of X were associated with higher values of Y, and if we also had extensive knowledge about the impact of confounders (e.g., mediating variables) across cases, this would then enable us in theory to estimate a propensity score for the likelihood of high values of X producing high values of Y in any particular case. Within medical research, personalized medicine has the ambition of collecting sufficient evidence of how treatments work across many different types of patients to enable estimation of propensity scores for individuals; it is arguable whether any IS research field has sufficient knowledge to be able to move from population-level evidence of mean causal effects (e.g., that high levels of intergroup inequality is associated with higher likelihoods of civil war) to any particular case. And given that claims about trends are typically being made across many cases, evidence of a causal relationship from a particular case would then tell us very little about the cross-case causal relationship (Humphreys and Jacobs 2015).

Actual experiments have further difficulty in communicating with case-based research because their inferences do not necessarily hold outside of controlled laboratory settings, meaning that the ability to infer to cases outside the lab is even further reduced (Deaton and Cartwright 2018). This is a particular concern as regards survey experiments, given that giving a respondent a particular message and then measuring their immediate response is far from the cacophonous world of real-world political communication faced by actual voters assessing foreign policy events.

There is an even more fundamental problem in making variance- and case-based designs communicate with each other. Variance-based designs provide evidence of a causal effect of X on Y by comparing across cases, but we do not learn what process links the two together—only that there is a mean causal effect (Dowe 2011; Illari 2011; Machamer 2004; Russo and Williamson 2007; Waskan 2011). An experiment does not tell us what links X and Y together, only that there is a mean causal effect across cases. In general, in variance-based designs, identifying causal mechanisms/processes that are operative in any given case is done using descriptive case studies. Weller and Barnes term this type of descriptive analysis pathway analysis *within* a single case, in which the analyst searches for indicators or observable manifestations that suggest the presence of a particular mechanism within a case. Weller and Barnes (2014, 21) argue that, “traditional guidelines for research design aimed at causal inference will likely be inappropriate.” Once a mechanism is identified, it is then treated as a counterfactual-based intervening variable that is given an empirical indicator that enables empirical assessment using robust, cross-case comparisons (e.g., mediation analysis).

The challenges of combining *case-based research* with variance-based findings are in many respects the inverse. First, taking individual cases as an analytical point of departure requires logically that we make ontologically *deterministic* causal claims about mechanisms (Mahoney 2008; Beach and Pedersen 2016, 19–24). Ontological determinism at the case-level means that an outcome occurred for a reason (or set

of reasons) in a given case. Ontological probabilism in a particular case would mean that outcomes can occur randomly at the case-level, which logically makes no sense. For instance, the UK decision to call for a referendum on EU membership was not a random event but was the product of a series of choices made by individual actors in a particular political context. Of course, we might not be able to *epistemically* identify the causes of a particular outcome in a case, but that does not mean it happened by chance.

The methodological consequence of ontologically deterministic causal claims is that we are not hedging our theoretical bets by making claims about trends but, instead, claims about how things work within particular cases. This implies that it can be very difficult to generalize from one case to others because of the contextual sensitivity of causal processes (Ragin 2000; Falletti and Lynch 2009; Goertz and Mahoney 2009; Gerring 2010). As with experimental designs in variance-based research, in case-based research we trade higher internal validity of our causal (mechanistic) inferences for a reduced ability to generalize beyond the studied setting (lower external validity). The trade-off can be expressed using Cartwright's language (2012). Case-based research tells us about how "it works here," but because of the level of detailed tracing of processes required to make strong causal inferences, it becomes difficult to generalize our findings to other cases because contextual conditions in other cases might impact how the process plays out. This can be termed the problem of *mechanistic heterogeneity*, in which either (1) the same cause triggers different mechanisms that are linked to different outcomes in different contexts, or (2) the same cause and outcome are linked together through different mechanisms in different contexts. While the first problem can be alleviated by only generalizing to other cases that share the same cause(s) and outcome, the second problem can only be alleviated by looking under the hood of other cases to detect whether similar processes were also operative there. This means, in effect, that if we want to generalize about mechanisms across cases, we need to engage in detailed process tracing of multiple cases in order to substantiate that the same process actually is operative across a set of cases.

Some scholars contend that we could lift the level of theoretical abstraction of our theorized mechanisms/processes to such a high level that contextual differences would wash out. As an example, Elster talks about context-free mechanisms using very abstract language such as "wishful thinking" or "the spillover effect," but this type of mechanistic theorization tells us nothing about the actual causal process and how it worked in real-world cases (1998, 45, 52–55). Therefore, a more productive way forward, instead of shifting to a very high level of abstraction, is to make more extensive claims about processes operative in smaller, bounded sets of cases (Bechtel and Richardson 2010). But appreciating complexity does not mean that we are trapped in making idiographic claims about particular cases. Indeed, we can engage in cumulative research through intensive collaborative research over a longer time period, producing an evidence-based catalogue of different mechanism mechanisms that are triggered by a given cause (or set of causes) in different contexts.

A final problem in making case-based tracing of mechanisms/processes and variance-based comparisons of mean causal effects communicate relates to the problem of "masking" from the philosophy of science (Steel 2008, 68; Clarke et al. 2014). Masking is defined as a situation in which a cause is linked to an outcome through *multiple* mechanisms that have different effects in the same case. The classic example is the relationship between exercise (X) and weight (Y). When one goes for a run, two different mechanisms are triggered. The first mechanism links X in a negative relationship with Y through a burning calories causal process. The second mechanism links X and Y in a positive relationship through a building muscle causal process. If we tested the relationship using variance-based methods, we might find that in a given experiment there was a positive net causal effect on

average across the studied subjects. In contrast, a case-based analysis might trace the burning calories mechanism and how it operates, but this evidence is uninformative about the net causal effect. When we are interested in mechanisms, we are typically answering a “how does it work?” question, whereas the variance-based approach would be answering an “it works (somewhere)?” type of question. To communicate more meaningfully together would require that the within-case analysis traces *all* of the mechanisms triggered by the cause (and its absence) *and* that the variance-based experiment is repeated for many different specifications of exercise (e.g., sprinting versus long-distance, lifting weights, etc.) and contexts (e.g., dietary variations) to investigate the impact of different variants of exercise under different contexts to figure out the context in which one of the mechanisms is more dominant and vice versa.

Interpretivist methodology is markedly different from both variance- and case-based research. To an interpretivist, the decision to employ experience-distant concepts to design an inquiry about a social phenomenon across several cases or within a single case constitutes what we denote as the *objectification fallacy*. Here we identify two key points of incommensurability. First, using experience-distant concepts to define a social phenomenon of interest objectively, variance- and case-based methodologies *dehumanize* social science by turning it into an inquiry of *things with essences* instead of *beings with existences*. To an interpretivist, this means that variance- and case-based methodologies from the onset fail to focus on what they believe ought to be the subject matter of social science—the existence of human beings as they experience it.

Second, failing to be aware of the objectification fallacy is not solely a core problem facing social science as a discipline. Lacking awareness of the implications of this fallacy prohibits social scientists from combatting stereotypes by disclosing otherwise “assumed, unspoken or taken-for-granted ideas” (Yanow and Schwartz-Shea 2012, 23). Combatting experience-distant stereotypes is important in social science because stereotypes—at best—are conceivable as problems about whether the findings mean what one thinks they do but can—at worst—lead to fatal policy decisions.<sup>8</sup>

#### *A Way Forward—Appreciating the Existence of Three Very Different Languages*

Concluding, the three methodologies ask fundamentally different questions that are evidenced using very different forms of empirical material. This means that claims and evidence from one methodology cannot travel easily across the divides. The only way to do this is to basically try to ignore the divide by transforming methods from different methodologies into pale shadows of themselves—unfortunately, an all too common strategy. This does not mean that using methods from other methodological approaches cannot be a useful research tool in combination with methods from one approach, only that we should explicitly acknowledge that they act as adjunct tools. For example, it can be helpful when designing a survey experiment to use more interpretive-inspired interviewing methods to figure out how respondents in a particular social context react and understand our survey questions. However, this preliminary interpretive probing would then be used pragmatically to improve the measurement validity of our survey questions. This “experience-distant” use of the interview data means that the interpretive-inspired interviews function as an adjunct method in relation to the core of the research, the variance-based survey experimental method.

<sup>8</sup> Fredric C. Schaffer (2014) problematizes generation and analysis of data about democracy relying on thin descriptions coming from the application of experience-distant concepts. Comparing the results from a 2002 survey about democracy conducted in the Philippines with his 2001 fieldwork in a Philippine community, Schaffer demonstrates that what democracy means to researchers studying democracy and to researched agents in a specific setting may differ significantly (Schaffer 2014, 327).

We have three suggestions to move methodological conversations forward. First, to avoid the compartmentalization of research into three distinct siloes in which scholars from one approach do not understand the claims and evidence produced by researchers from other approaches, we suggest that all scholars should be trained in the basic workings and underlying foundations of all three research methodologies. Ideally, this will take place early in the training of scholars to avoid them becoming cemented in the view that only one methodology is “scientific.” At our department at Aarhus University, we have a core mandatory PhD methods course at the start of the program, in which participants are taught about the basic ontological and epistemological assumptions that underlie the three different methodologies. By the end of the course, participants have learned to appreciate the differences, enabling them to debate for instance the benefits and trade-offs from investigating a particular topic using either an interpretive or variance-based design. The goal is not to make participants fluent in methods within each of the methodologies, but for them to understand and appreciate the core *differences* across the methodological divides.

Second, we suggest that scholars then specialize in methods within one methodology, given the increasing sophistication of methods within each of the three methodologies. This specialization should focus in particular on understanding the fundamental ontological and epistemological foundations of the approach within which they operate, the research implications that these have, and the core points of contention *within* the tradition.

There has luckily been an increased recognition for the need for clarity about these foundations in different methodologies. In case-based research, there is increasing recognition of the fundamental differences between within-case and cross-case analysis (e.g., Goertz and Mahoney 2012; Beach and Pedersen 2016). Within interpretivism, there is an increasing recognition of the need to further develop a parallel scientific language to conduct and communicate about human meaning-making on interpretivist terms, instead of using variance- or case-based terminology (e.g., Bevir and Rhodes 2010, 2016; Yanow and Schwartz-Shea 2012, 2014). In variance-based research, the potential outcomes framework is now widely understood to be at the core of the approach.

Finally, we do not suggest that multimethodological research is completely impossible but that it will typically require collaboration between scholars fluent in different methodologies. Therefore, we suggest that multimethodological research be undertaken by teams of scholars who have specialized in their respective methodologies. The goal though would not be to triangulate because of the fundamental differences but instead to gain different insights about a shared research theme. This could, for instance, involve a set of variance-based studies of public opinion using experiments and large-N surveys, combined with interpretivist focus groups within particular social contexts. The findings would not seamlessly communicate with each other because of the difficulties in translating claims and evidence across the divides but would both provide important differing insights about the social world that we all care so much about.

## References

- ABBOTT, ANDREW. 1995. “Things of Boundaries.” *Social Research* 62 (4): 857–82.
- ABELL, PETER. 2004. “Narrative Explanation: An Alternative to Variable-Centered Explanation?” *Annual Review of Sociology* 30 (1): 287–310.
- AHMED, AMEL, AND RUDRA SIL. 2009. “Is Multi-method Research Really ‘Better’?” *Qualitative and Multi-Method Research* 7 (2): 2–6.
- AHN, ANDREW C., MUNEESH TEWARI, CHI-SANG POON, AND RUSSELL S PHILLIPS. 2006. “The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative?” *PLoS Medicine* 3 (6): e208. <https://doi.org/10.1371/journal.pmed.0030208>



- ANGRIST, JOSHUA A., AND JÖRN-STEFFEN PISCHKE. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- ANScombe, GERTRUDE E. M. 1957. *Intention*. Oxford: Basil Blackwell.
- BEACH, DEREK, AND RASMUS BRUN PEDERSEN. 2016. *Causal Case Studies: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor: University of Michigan Press.
- . 2019. *Process-Tracing Methods*. 2nd ed. Ann Arbor: University of Michigan Press.
- BEACH, DEREK, AND INGO ROHLFING. 2018. "Integrating Cross-Case Analyses and Process Tracing in Set Theoretic Research: Strategies and Parameters of Debate." *Sociological Methods and Research* 47 (1): 3–36.
- BECHTEL, WILLIAM, AND ROBERT C. RICHARDSON. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.
- BENNETT, ANDREW, AND JEFFREY CHECKEL. Eds. 2014. *Process Tracing: From Metaphor to Analytical Tool*. Cambridge: Cambridge University Press.
- BERGER, PETER L., AND THOMAS LUCKMANN. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books.
- BEVIR, MARK, AND JASON BLAKELY. 2018. *Interpretive Social Science: An Anti-naturalist Approach*. Oxford: Oxford University Press.
- BEVIR, MARK, AND ASAF KEDAR. 2008. "Concept Formation in Political Science: An Anti-naturalist Critique of Qualitative." *Perspectives on Politics* 6 (3): 503–17.
- BEVIR, MARK, AND RODERICK A. W. RHODES. 2010. *The State as Cultural Practice*. Oxford: Oxford University Press.
- . 2016. *Routledge Handbook of Interpretive Political Science*. London: Routledge.
- BHASKAR, ROY. 1978. *A Realist Theory of Science*. Brighton: Harvester.
- BOGEN, JIM. 2005. "Regularities and Causality: Generalizations and Causal Explanations." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 397–420.
- BRADY, HENRY E., AND DAVID COLLIER. 2011. *Rethinking Social Inquiry: Diverse Tools Shared Standards*. 2nd ed. Lanham, MD: Rowman Littlefield.
- BRULÉ, DAVID J. 2008. "Congress, Presidential Approval, and U.S. Dispute Initiation." *Foreign Policy Analysis* 4 (4): 349–70.
- BYGSTAD, BENDIK, AND BJØRN ERIK MUNKVOLD. 2011. "Exploring the Role of Informants in Interpretive Case Study Research in IS." *Journal of Information Technology* 26 (1): 32–45.
- CARR, EDWARD H. 2001. *What Is History?* Houndmills, UK: Palgrave.
- CARTWRIGHT, NANCY. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.
- . 2011. "Predicting 'It Will Work for Us': (Way) beyond Statistics." In *Causality in the Sciences*, edited by Phyllis McKay Illari, Federica Russo and Jon Williamson, 750–68. Oxford: Oxford University Press.
- . 2012. "Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79 (5): 973–89.
- CHATTERJEE, ABHISHEK. 2011. "Ontology, Epistemology, and Multimethod Research in Political Science." *Philosophy of the Social Sciences* 43 (1): 73–99.
- CLARKE, B., D. GILLIES, PHYLLIS ILLARI, FEDERICA RUSSO, AND JON WILLIAMSON. 2014. "Mechanisms and the Evidence Hierarchy." *Topoi* 33 (2): 339–60.
- DAFOE, ALLAN, JOHN R. ONEAL, AND BRUCE RUSSETT. 2013. "The Democratic Peace: Weighing the Evidence and Cautious Inference." *International Studies Quarterly* 57 (1): 201–14.
- DEATON, ANGUS, AND NANCY CARTWRIGHT. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210 (2018): 2–21.
- DOWE, PHIL. 2011. "The Causal-Process-Model Theory of Mechanisms." In *Causality in the Sciences*, edited by Phyllis McKay Illari, Federica Russo and Jon Williamson, 865–79. Oxford: Oxford University Press.
- DUNNING, THAD. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- EMIRBAYER, MUSTAFA. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103 (2): 281–317.
- FAIRFIELD, TASHA. 2019. "A Dialogue with the Data: The Bayesian Foundations of Iterative Research in Qualitative Social Science." *Perspectives on Politics* 17 (1): 154–67.
- FALLETTI, TULIA G., AND JULIA F. LYNCH. 2009. "Context and Causal Mechanisms in Political Analysis." *Comparative Political Studies* 42 (9): 1143–66.
- FEARON, JAMES. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43 (2): 169–95.
- FEARON, JAMES, AND ALEXANDER WENDT. 2005. "Rationalism V. Constructivism: A Skeptical View." In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse and Beth A. Simmons, 52–72. London: SAGE.

- FUJII, LEE ANN. 2008. "The Power of Local Ties: Popular Participation in the Rwandan Genocide." *Security Studies* 17 (3): 568–97.
- . 2017. *Interviewing in Social Science Research: A Relational Approach*. London: Routledge.
- GEERTZ, CLIFFORD. 1973. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.
- . 1974. "From the Native's Point of View: On the Nature of Anthropological Understanding." *Bulletin of the American Academy of Arts and Sciences* 28 (1): 26–45.
- GEORGE, ALEXANDER L., AND ANDREW BENNETT. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- GERRING, JOHN. 2010. "Causal Mechanisms: Yes But . . ." *Comparative Political Studies* 43 (11): 1499–526.
- . 2011. *Social Science Methodology—A Unified Framework*. Cambridge: Cambridge University Press.
- GOERTZ, GARY. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton, NJ: Princeton University Press.
- GOERTZ, GARY, AND JACK S. LEVY, eds. 2007. *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*. London: Routledge.
- GOERTZ, GARY, AND JAMES MAHONEY. 2009. "Scope in Case-Study Research." In *The Sage Handbook of Case-Based Methods*, edited by David Byrne and Charles C. Ragin, 307–17. Thousand Oaks, CA: SAGE.
- . 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- GROFF, RUTH. 2011. "Getting Past Hume in the Philosophy of Social Science." In *Causality in the Sciences*, edited by Phyllis McKay Illari, Federica Russo and Jon Williamson, 296–316. Oxford: Oxford University Press.
- HANSEN, LENE. 2006. *Security as Practice: Discourse Analysis and the Bosnian War*. New York: Routledge.
- HOLLAND, PAUL W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.
- HOPF, TED. 1998. "The Promise of Constructivism in International Relations Theory." *International Security* 23 (1): 171–200.
- . 2002. *Social Construction of International Politics: Identities & Foreign Policies, Moscow, 1955 and 1999*. Ithaca, NY: Cornell University Press.
- HUMPHREYS, MACARTAN, AND ALAN JACOBS. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109 (4): 653–73.
- ILLARI, PHYLLIS MCKAY. 2011. "Mechanistic Evidence: Disambiguating the Russo-Williamson Thesis." *International Studies in the Philosophy of Science* 25 (2): 139–57.
- ILLARI, PHYLLIS MCKAY, AND JON WILLIAMSON. 2011. "Mechanisms Are Real and Local." In *Causality in the Sciences*, edited by Phyllis McKay Illari, Federica Russo and Jon Williamson, 818–44. Oxford: Oxford University Press.
- JACKSON, PATRICK THADDEUS. 2006. *Civilizing the Enemy: German Reconstruction and the Invention of the West*. Ann Arbor: University of Michigan Press.
- . 2014. "Making Sense of Making Sense: Configurational Analysis and the Double Hermeneutic." In Yanow and Schwartz-Shea, *Interpretation and Method*, 267–83.
- . 2016. *The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics*. 2nd ed. London: Routledge.
- JACKSON, PATRICK THADDEUS, AND DANIEL H. NEXON. 1999. "Relations before States: Substance, Process and the Study of World Politics." *European Journal of International Relations* 5 (3): 291–332.
- JUNG, HOYOUN. 2019. "The Evolution of Social Constructivism in Political Science: Past to Present." *SAGE Open*. (January–March): 1–10. <https://doi.org/10.1177/2158244019832703>.
- KHONG, YUEN FOONG. 1992. *Analogies at War: Korea Munich Dien Bien Phu and the Vietnam Decisions of 1965*. Princeton: Princeton University Press.
- KREBS, RONALD R., AND JENNIFER K. LOBASZ. 2007. "Fixing the Meaning of 9/11: Hegemony, Coercion, and the Road to War in Iraq." *Security Studies* 16 (3): 409–51.
- KHOSROWI, DONAL. 2019. "Extrapolation of Causal Effects: Hopes, Assumptions, and the Extrapolator's Circle." *Journal of Economic Methodology* 26 (1): 45–58. <https://doi.org/10.1080/1350178X.2018.1561078>.
- KING, GARY, ROBERT O. KEOHANE, AND SIDNEY VERBA. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- KREUZER, MARKUS. 2020. *The Grammar of Time: Using Comparative Historical Analysis to Study the Past*. Cambridge: Cambridge University Press.
- LEAMER, EDWARD E. 2010. "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives* 24 (2): 31–46.
- LEBOW, RICHARD NED. 2000–2001. "Contingency, Catalysts and International System Change." *Political Science Quarterly* 115, no. 4 (Winter): 591–616.

- LEVI-MONTALCINI, R., AND P. CALISSANO. 2006. "The Scientific Challenge of the 21st Century: From a Reductionist to a Holistic Approach via Systems Biology." *BMC Neuroscience* 7 (Supplement 1): S1. <https://doi.org/10.1186/1471-2202-7-S1-S1>.
- LEWIS, D. 1986. *Causation: Postscripts to "Causation."* Philosophical papers, vol. II. Oxford: Oxford University Press.
- LEVY, JACK. 2015. "Counterfactuals, Causal Inference, and Historical Analysis." *Security Studies* 24 (3): 378–402.
- LIEBERMAN, EVAN S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–51.
- LIEBERSON, STANLEY. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70 (2): 307–20.
- LYNCH, CECILIA. 1999. *Beyond Appeasement: Interpreting Interwar Peace Movements in World Politics*. Ithaca, NY: Cornell University Press.
- MACHAMER, PETER. 2004. "Activities and Causation: The Metaphysics and Epistemology of Mechanisms." *International Studies in the Philosophy of Science* 18 (1): 27–39.
- MACHAMER, PETER, LINDLEY DARDEN, AND CARL F. CRAVER. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.
- MAHONEY, JAMES. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41 (4–5): 412–36.
- MCCOURT, DAVID M. 2016. "Practice Theory and Relationalism as the New Constructivism." *International Studies Quarterly* 60 (3): 475–85.
- MORGAN, STEPHEN L., AND CHRISTOPHER WINSHIP. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- MORISON, SAMUEL ELIOT. 1951. "Faith of a Historian." *The American Historical Review* 56 (2): 261–75.
- ONUF, NICOLAS GREENWORD. 1989. *World of Our Making: Rules and Rule in Social Theory and International Relations*. Columbia: University of South Carolina Press.
- OREN, IDO. 1995. "The Subjectivity of the 'Democratic' Peace: Changing U.S. Perceptions of Imperial Germany." *International Security* 20 (2): 147–84.
- PEDERSEN, JONAS GEJL. 2019. *The Russian Quest for Ontological Security*. Aarhus, Denmark: Politica.
- QIN, YAQING. 2018. *A Relational Theory of World Politics*. Cambridge: Cambridge University Press.
- RAGIN, CHARLES C. 1987. *The Comparative Method Moving beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- . 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- RUBIN, DONALD. 1980. "Discussion of 'Randomization Analysis of Experimental Data in the Fisher Randomization Test' by Basu." *Journal of the American Statistical Association* 75 (371): 591–93.
- RUBIN, DONALD. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100 (469): 322–31.
- RUSSO, FEDERICA, AND JON WILLIAMSON. 2007. "Interpreting Causality in the Health Science." *International Studies in the Philosophy of Science* 21 (2): 157–70.
- . 2011. "Generic versus Single-Case Causality: The Case of Autopsy." *European Journal of the Philosophy of Science* 1 (1): 47–69.
- SAYER, ANDREW. 2000. *Realism and Social Science*. London: Sage.
- SCHAFER, FREDERIC CHARLES. 2014. "Thin Descriptions: The Limits of Survey Research on the Meaning of Democracy." *Polity* 46 (3): 303–30.
- . 2016. *Elucidating Social Science Concepts: An Interpretivist Guide*. New York: Routledge.
- SCHNEIDER, CARSTEN Q., AND INGO ROHLFING. 2013. "Combining QCA and Process Tracing in Set-Theoretical Multi-method Research." *Sociological Methods and Research* 42 (4): 559–97.
- . 2016. "Case Studies Nested in Fuzzy-Set QCA on Sufficiency: Formalizing Case Selection and Causal Inference." *Sociological Methods and Research* 45 (3): 526–68.
- SCHWARTZ-SHEA, PEREGRINE. 2014. "Judging Quality: Evaluative Criteria and Epistemic Communities." In Yanow and Schwartz-Shea, *Interpretation and Method*, 120–46.
- SEAWRIGHT, JASON. 2016. *Multi-method Social Science*. Cambridge: Cambridge University Press.
- TEEL, DANIEL. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.
- TAVORY, IDDO, AND STEFAN TIMMERMANS. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.
- TAYLOR, CHARLES. 1971. "Interpretation and the Sciences of Man." *The Review of Metaphysics* 25 (1): 3–51.
- TETLOCK, PHILIP E., AND AARON BELKIN, eds. 1996. *Counterfactual Thought Experiments in World Politics. Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press.

- THELEN, KATHLEEN,, AND JAMES MAHONEY. 2015. "Comparative-Historical Analysis in Contemporary Political Science." In *Advances in Comparative-Historical Analysis*, edited by Mahoney and Thelen, 3–38. Cambridge: Cambridge University Press.
- TOMZ, MICHAEL R., AND JESSICA L. P. WEEKS. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107 (4): 849–65.
- WASKAN, JONATHAN. 2011. "Mechanistic Explanation at the Limit." *Synthèse* 183 (3): 389–408.
- WELDES, JUTTA. 1999. *Constructing National Interests: The United States and the Cuban Missile Crisis*. Minneapolis: University of Minnesota Press.
- WELLER, NICHOLAS, AND JEB BARNES. 2014. *Finding Pathways: Mixed-Method Research for Studying Causal Mechanisms*. Cambridge: Cambridge University Press.
- WENDT, ALEXANDER. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- WILLIAMS, MALCOLM, AND WENDY DYER. 2009. "Single Case Probabilities." In *Case Based Methods*, edited by David Byrne and Charles C. Ragin, 84–100. London: SAGE.
- WOODWARD, JAMES. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- YANOW, DVORA, AND PEREGRINE SCHWARTZ-SHEA. 2012. *Interpretive Research Design: Concepts and Processes*. New York: Routledge.
- . 2014. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. 2nd ed., edited by Dvora Yanow and Peregrine Schwartz-Shea. New York: M.E. Sharpe.