



AARHUS UNIVERSITY



Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Beach, D., & Kaas, J. G. (2020). The Great Divides: Incommensurability, the Impossibility of Mixed-Methodology, and What to Do about It. *International Studies Review*, 22(2), 214-235.

<https://doi.org/10.1093/isr/viaa016>

Publication metadata

Title:	The Great Divides: Incommensurability, the Impossibility of Mixed-Methodology, and What to Do about It
Author(s):	Derek Beach & Jonas Gejl Kaas
Journal:	<i>International Studies Review</i> , 22(2), 214-235
DOI/Link:	https://doi.org/10.1093/isr/viaa016
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The great divides: incommensurability, the impossibility of mixed-methodology, and what to do about it

Derek Beach & Jonas Gejl Kaas

Department of Political Science, Aarhus University

Abstract

There is still significant confusion about how multi-method 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 *multi-method* and *multi-methodology* 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 aim to learn 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 relations scholarship, and to provide suggestions for what we can do about it.

Keywords:

Multi-method research, methodology, case-based, variance-based, causal mechanisms, counterfactuals, interpretivism, meaning-making

Multi-method research has become increasingly popular in recent years in the study of international relations (IR) (Lieberman, 2005; Schneider and Rohlfing, 2013, 2016; Humphrey and Jacobs, 2015; Seawright, 2016; Goertz, 2017; Beach and Rohlfing, 2018).¹ The most common combination involves some form of cross-case comparative analysis (e.g. statistically assessing mean causal effects of a large number of cases) and in-depth case analysis (e.g. using process-tracing case studies). The promise of multi-method research is that different methods can compensate for each other's relative weaknesses, enabling more robust inferences to be made. There is even a widespread perception amongst many scholars that high-quality book length work *has* to engage in multi-method research on a topic.

Yet there is still considerable confusion about how multi-method research can be undertaken, and even if it is possible. In this article, we contend that much of the confusion is the result of a failure to distinguish between *multi-method* and *multi-methodology* 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 broader philosophy of science literature, 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 basis 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 claim that are evidenced using observational traces (aka mechanistic evidence), often supplemented with bounded 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

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

unhelpful term that tends to conflate two approaches with fundamentally different ontological and epistemological foundations (Yanow and Schwartz-Shea, 2012, 2014).

Multi-method 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, multi-methodological research would involve combining methods from *different* methodologies, for example by 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 multi-method research is possible, multi-methodology 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 method would be to investigate the co-variation 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 would select a particular case of an outbreak of intrastate war and for instance 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 life 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 co-variational 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 multi-methodology 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 (in)famously seen in King, Keohane and Verba's *Designing Social Inquiry*, in which they claim there is only one logic of social inquiry. They then contend that an interpretation of the ontological foundations of variance-based research of regularity and counterfactuals should be imposed on case study methods by transforming them into quasi-experimental assessments of difference-making across cases.² Monism is also found amongst proponents of other methodologies, for example witnessed in Bevir and Blakely's claim in *Interpretive Social Science* that, 'anyone who really desires to understand and explain human agency must adopt a 'hermeneutic' or interpretive philosophical perspective.' (2018: 1), in effect arguing for shoehorning methods from other methodologies into an interpretive world.

In monist multi-method research, methods from other methodologies therefore become adjunct tools used to merely illustrate claims that are more strongly evidenced using methods 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 studies of particular cases that are merely *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 fundamental assumptions that make it difficult for them to communicate with each other, and that make it impossible to claim that they

² - For similar arguments, see Gerring, 2011. For critical discussions from a case-based side, see Ragin, 2000; Brady and Collier, 2010; Goertz and Mahoney, 2012; Beach and Pedersen, 2016.

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 IR.

This article starts by introducing the ontological and epistemological underpinnings of different methodologies by differentiating into a variance-based, case-based, and an interpretivist approach. For each, we evaluate the epistemological consequences that differing fundamental ontological assumptions have and provide examples of IR 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.

2. Variance-based methodology - 'It works somewhere' claims

Variance-based methodology is a 'top-down' form of research, in which the focus is on assessing mean causal effects across a set of cases. In variance-based approaches, the evidential gold standard is randomized controlled experiment, which if properly designed to ensure strong control for potential confounders, enables strong *causal* inferences 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 (confounders) due to the controlled setting (Gerring, 2011; Clarke et al, 2014).³

Strictly speaking, the level at which causes are operative is always *within* a single case. 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 war-making choices of other Presidents unless there are important learning effects in subsequent cases that link cases together over time. In other words, 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. Yet while causation *always* takes place *within* cases, in variance-based research we *learn* about causation by studying cross-case variation. 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.

Variance-based approaches build on the *counterfactual* understanding of causation – often developed as the potential outcomes framework - situated firmly within a neopositivist understanding of science (Woodward, 2003; Rubin, 2005; Angrist and Pischke, 2009; Jackson, 2016). Counterfactual causation is defined as the claim that a cause produced an outcome because its absence would result in the absence of the outcome, all other things being held equal (Lewis 1986: 160; Woodward 2003). Without evaluating the *difference* that a cause can make between the actual case and the counterfactual non-treated case in a controlled setting, no *causal* inferences are possible.

³ - Within medicine, the top of the hierarchy is actually meta-studies that evaluate the findings of multiple RCTs on the same research question in different settings.

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 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 is impossible without gaining access to what quantum physics terms the multiverse of infinite potential parallel counterfactual worlds, 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. This is best seen in a controlled experiment, where we compare values of the outcome in cases that receive the treatment with those in the experimental control group that do not (i.e. the counterfactual state), holding other factors constant. Here the lack of treatment in the control group acts as the counterfactual, enabling us to infer that if there is a significant and substantial difference in values of the outcome in the two groups, this difference is the mean causal effect of the treatment. Given the need to compare *across* cases, variance-based approaches can be termed a ‘top-down’ form of research (Illari and Williamson, 2011). Again, this is best seen in an experiment, where mean causal effects (the *average* ‘difference’ that the cause makes for the outcome across the treatment and control groups) are assessed within the population of cases in the study. The term top-down is therefore appropriate because causation is studied at the population level (or samples thereof) by assessing trends of differences *across* cases.

Key to our ability to make causal inferences are the assumptions of unit homogeneity and independence of units (Holland, 1986; King, Keohane and Verba, 1994: 91-97). *Unit homogeneity* means that the same cause will produce the same results in two or more cases. *Independence of units* means that the potential outcome in one case is

unaffected by values of the cause in other cases. If these two assumptions do not hold, we will have biased estimates of the difference that variations in X have for values of Y. 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 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 (i.e. *internal validity*) (Gerring, 2011; Clarke et al, 2014). Controlled experimental designs in the form of lab, survey or field experiments are at the top of the hierarchy because the random assignment of a large number of cases into treatment and control groups ensures that the units are on average homogeneous, thereby ensuring that any differences between units wash 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 US and UK in which they asked a large number of respondents whether they would support a pre-emptive strike on a country that could develop nuclear weapons within six months. The experimental manipulation related to giving 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 IR 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, ‘...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.’ (2013: 202). Here it is the ability to control for potential confounders across a large number of cases that enables cautious causal inferences to be made.

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 IR 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 (Lieberson, 1991). Ideally, we would be able to identify two cases that are similar in all respects but the independent variable of interest. But as Runhardt (2015: 1306) admits, ‘A similarity comparison in areas like political science is, however, difficult to defend.’ Because of the complexity of IR phenomena, it is difficult to find cases in which the ‘all other

things equal' assumption required for causal inferences actually holds (Ragin 1987: 48). Levy (2015: 390) writes that 'Controlled comparison and matching face the difficulty of finding real-world cases that are identical in all respects but one.' But unless we can substantiate that all other things are equal except for the presence/absence of a cause, we cannot make a causal inference that its absence made a difference for the outcome.

At the 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 approach need to be transformed into multi-case comparisons. This can be achieved by the strategy suggested by King, Keohane and Verba (1994: 217-228), where they argue that we should transform 'one case into many' by disaggregating a case either spatially or temporally, enabling a (weak) assessment of the counterfactual in the form of a most-similar-system comparison (everything else is equal except variation in the cause) (King, Keohane and Verba, 1994: 217-228). However, this creates a situation where our 'cases' are neither homogeneous nor independent. For example, disaggregating an international negotiation as a case into temporal stages ($t_0, t_1, \dots t_n$) results in cases that are quite causally dissimilar because we should expect critical differences in how causes play out when comparing early stages (agenda-setting) and the end game. In addition, the 'cases' would not be independent of each other because in a negotiation what happens at the start (t_0) naturally affects events later in the negotiation, meaning that values of Y in case t_0 will influence values of X in subsequent cases. King, Keohane, and Verba (1994: 222) even admit that this is a problem, concluding, 'When dealing with partially dependent observations, we should be careful not to overstate the certainty of the conclusions.'

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. The logical argument is then made that if a particular cause had not occurred, the outcome would not have

occurred (Goertz and Levy 2007; Tetlock and Belkin 1996; Lebow 2000; Levy 2015 Fearon 1991). However, the evidence is purely hypothetical, and we have no way of knowing whether things really would have been different if the cause was not present.

Therefore, in variance-based approaches, case studies can at best help us detect measurement error or potential confounders when we engage in more exploratory research that can help us improve the statistical models we use to explore population-wide difference-making (Seawright, 2016: 45-69). More often, variance-based multi-method designs use case studies as mere descriptive illustrations, with the core causal inferences based on the large-n controlled comparison.

Concluding, variance-based approaches are top-down methods that assess counterfactual causation in the form of mean causal effects across cases. Relative strengths include the ability to assess the magnitude of net causal effects, and the ability to make causal inferences about many cases (populations or samples thereof) when we are able to deploy actual or natural experiments. However, 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 a trade-off between the *internal* validity and *external* (i.e. ability to infer to other cases outside of the studied cases) validity. Lab or survey experiments have a high level of internal validity, but they typically have low external 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 US in the 1960s, or other democratic countries that have very different political systems (e.g. Japan)) (i.e. low external validity). Natural experiments have lower internal validity because we are *assuming* that the situation approximates an experiment. 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

validity, whereas small-n comparisons and single case studies have neither internal nor external validity.

3. Case-based methodology - how causes work in cases

The core of case-based methodology is *within-case* tracing of causal processes (also termed mechanisms)⁴ using detailed case studies, although cross-case comparisons are important for selecting appropriate cases and generalizing findings about processes within 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 exploring the similarities and differences in how the subsequent process plays out in the cases (see e.g. Kruezer forthcoming; Thelen and Mahoney, 2015).

Building on critical realist foundations (e.g. Bhaskar, 1978; Salmon, 1998; Jackson, 2016), case-based methodology is a form of ‘bottom-up’ research because the in-depth study of individual cases is the analytical point of departure. Here the central goal is to learn about causal processes and how they play out in particular cases (Russo and Williamson, 2011). This means also that case-based research is typically very sensitive to the *context* in which causal processes operate (Thelen and Mahoney, 2015: 7-15; Ragin, 2000: 53-63; Falletti and Lynch 2009). Scholars within the case-based approach claim that complexity and contextual sensitivity of causal relationships are key features of 21st-century science, seen in developments in fields like systems biology or personalized medicine (Ahn et al. 2006; Bechtel and Richardson 2010; Cartwright 2007, 2012; Khosrowi, 2019; Levi-Montalcini and Calissano 2006).

⁴ - We use the terms causal process and causal mechanisms interchangeably in this section.

Causal processes in case-based research are not causes; they are what link causes and outcomes together. In a case-based understanding, processes/mechanisms are therefore also more than just lower-level counterfactual claims. If one takes process seriously, the analytical goal is to explore the process that *actually* was operative *within* a case (Groff, 2011; Waskan, 2011; Machamer, 2004: 31). A ‘mechanism explanation for some happening that perplexes us is explanatory precisely in virtue of its capacity to enable us to understand how the parts of some system actually conspire to produce that happening’ (Waskan 2011: 393). 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?’. 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. The essence of mechanistic/process-based explanations is that we shift the analytical focus from causes and outcomes to what links them together in a productive relationship.

In case-based approaches, the core elements of a causal process are unpacked theoretically and studied empirically in the form of the traces left by the activities associated with each part of the process. Each of the parts of the mechanism/process can be described in terms of entities that engage in activities (Machamer 2004; Machamer, Darden, and Craver 2000). Entities are the factors (actors, organizations or structures) engaging in activities, whereas the activities are the producers of change or what transmits causal forces or powers through a mechanism. Mechanisms are therefore viewed in case-based research in a more holistic fashion than being mere simple counterfactuals, meaning that the effects of a mechanism/process are more than the sum of its parts. When a causal process is unpacked theoretically, the goal becomes to understand *how a process actually works* by tracing the operation of each part (or at least the most critical parts) in one or more cases. Exploring how a causal process worked within a case is often a back-and-forth, abductive research approach in which one's initial ideas about how it worked are empirically incorrect (Tavory and Timmermans, 2014). This leads to a phase of theory-building, trying to

understand how the process played out that can subsequently be tested again using new mechanistic evidence.

Mechanisms/processes are traced empirically by collecting what can be termed *mechanistic evidence*, defined as the observable fingerprints left by the operation of the activities associated with parts of mechanisms (Russo and Williamson, 2007; Illari, 2011). Here there is *no variation across cases*; instead it is the empirical traces and their association with activities left within a particular case that enable us to infer that we have evidence of a mechanism linking a cause (or set of causes) with an outcome (Clarke et al. 2014; Beach and Pedersen, 2019). Mechanistic evidence is observational data that enables causal inferences within cases to be made. In this respect, the fundamental problem of causal inference is therefore only a problem for variance-based approaches. While evidencing a counterfactual claim requires assessing 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 of processes using mechanistic evidence within individual cases is at the top of the evidential hierarchy as regards the internal validity of causal claims. 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 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.

Below this are weaker within-case methods that only obliquely trace mechanisms (congruence studies and analytical narratives), thereby not enabling strong causal inferences (e.g. George and Bennett, 2005; Abell, 2004). 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 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 works 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.⁵ 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

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

the low external validity might seem to be a critical weakness, the contextual sensitivity of claims in case-based research is seen by proponents of case-based designs as a critical strength (e.g. Ragin 2000; 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 also. 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.

4. 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, p. 23). Whereas interpretivist researchers debate whether interpretivism reflects an explicitly anti-naturalist stance (Bevir and Blakely, 2018), we see a recurring key aspiration among interpretivist researchers to elucidate context-specific meanings 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, p. 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. Scholars such as Mark Bevir, Roderick A.W. Rhodes, Dvora Yanow and Peregrine Schwartz-Shea’s works 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).⁶

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 IR (e.g. Jackson and Nexon, 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 differs from prominent IR constructivist Alexander Wendt's (1999) systemic theory of three cultures of anarchy (Hobbesian, Lockean and Kantian). Whereas Wendt argues 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 (ibid).

IR scholars interested in human meaning making often have constructivism as their theoretical point of departure. However, not all IR constructivists are interpretivists (e.g. 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 IR there are theoretical constructivists (e.g. (Hopf, 1998; Jung, 2019), and those who hold constructivist ontological views. 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 ethnography to elucidate the contestations and commonplaces of meanings concerning the post-WWII reconstruction of “Western Civilization” and implications

⁶ Contributions by Mark Bevir, Roderick A.W. Rhodes, Dvora Yanow and Peregrine Schwartz-Shea build on top of pioneering social science interpretivist work as *The Social Construction of Reality* (Berger and Luckmann, 1966), (1966), ‘Interpretation and the Sciences of Man’ (Taylor, 1971) and *The Interpretation of Cultures* (Geertz, 1973).

for transatlantic ties in-between German politicians and scholars. Theoretically, Hopf and Jackson both draw on social constructivism within IR, 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 vacuum, but 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, interpretivist 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, p. 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, p. 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 non-mechanistic 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, p. 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, p. 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, p. 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 reflects a general negligence of researchers’ will and capacity to reflect upon their positionality and the ideals, values and power structures influencing their construction and use of social science concepts. Oren

(1995: 178) concludes, 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 an 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, pp. 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).

Interpretivist 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, p. 85). Dvora Yanow notes, 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, p. 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 worthwhile 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 are not unproblematic since the trustworthiness of knowledge-claims rely on the researcher making these (Yanow and Schwartz-Shea, 2014, p. xxi; Pedersen, 2019, pp. 133–134 & 141).

Returning to researcher's will and ability to critically reflect on exposure, we find what Samuel E. Morison denotes *mesure* relevant. Morison (1951) defines *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 'becoming humility toward his fellow workers, and skepticism directed toward himself as toward them, will be of assistance' may help.

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 ‘becoming humility’, a good general starting point for producing thick descriptions is *listening* (Pedersen, 2019, p. 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 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, pp. 27–34, 2014, pp. 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 into researched setting, the researcher becomes able to better make sense out of agents’ meaning-making. Unlike variance-based approaches, interpretivist research designs are intentionally flexible to accommodate the understandings about researched that gradually emerges as 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, maintaining the researcher as simultaneously ‘being a stranger and being familiar’ to a researched site and agents is central to producing valid interpretivist knowledge (Yanow and Schwartz-Shea, 2012, p. 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, p. 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 setting. Having cultivated the researcher to researched, research moves on to encounter between researcher and researched. Encounter constitutes an iterative process of observing and reflecting. After encounter comes writing. The core aim of the writing is becoming aware of what researcher knows and—more importantly—what researcher still needs to know. Edward H. Carr encapsulated the aim and process of (re-)writing nothing 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, p. 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 evidence such descriptions become as support for 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 non-science (Jackson, 2016, p. 209). Besides enables public criticism (Jackson, 2016, p. 209), listening to comments and reviews from peers as well as members of the researched setting⁷ sharpens the accuracy and trustworthiness of presented knowledge-claims.

5. 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.

Variance-based	Case-based	Interpretive
----------------	------------	--------------

⁷ Frequently denoted as member-checking (Yanow and Schwartz-Shea, 2012, pp. 106–107) or informant feedback (Bygstad and Munkvold, 2011).

Analytical point of departure	Population-level (top-down)	Case-level (bottom-up)	Individuals (meaning making in context)
Types of claims	Average 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 + Mill's methods/QCA	Any method that enables trustworthy thick description, e.g. textual and ethnographic methods

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

Why are the methodologies incommensurable?

In this section, we discuss why the three methodologies are 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. Nancy Cartwright (2011) has succinctly defined the essence of the evidence that robust variance-based methods produce; ‘it works somewhere’. This means that we have ontologically *probabilistic* claims about *mean* causal effects in a population (or a sample thereof). However, evidence of a trend makes it almost impossible to communicate meaningfully with claims about single instances, 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 ignore the further difficulties of communicating with findings about relational meaning-making, it is very difficult to move to the level of individual cases because of the risk of causal heterogeneity. If a population was completely causally homogeneous, unit homogeneity would hold perfectly (Morgan and Winship, 2007: 37-40; Rubin, 1980: 961), meaning that population-level trends would also be perfectly predictive for effects in individual cases (Cartwright, 2009: 154-159). But given the causal complexity of the real world, there can be many reasons that the relationship does not hold in individual cases, including omitted variables such as contextual factors (Williams and Dyer 2009: 210-211). This is the very reason that ontologically probabilistic claims are made about trends in variance-based research.

When one then moves from population-level causal claims about trends to individual cases, causes are seen as ‘probability-raisers’ (Gerring, 2011: 199). Mean causal effects are averages *across* a study population, and there can be many ways in which causes and contextual factors combine to produce the same outcome in a complex, real-world population (Cartwright, 2012: 980-981; Leamer, 2010). In theory, if there is a positive relationship between X and Y, a high value of X would make it more probable that we would find a high value of Y in a particular case. We could then use what we know about mean causal effects of different independent variables and the impact of confounders, enabling us in theory to estimate propensity scores for individual cases. However, to do this requires either that we have evidence of a high

level of causal homogeneity in the population being studied that enables one to *assume* overall treatment effects apply to individual cases in a predictable fashion, or we have in effect mapped the causal heterogeneity embedded within the population, enabling cases to be grouped together into more homogeneous sub-sets of cases (e.g. there is a negative relationship between X and Y when factor Z1 is present, whereas there is no relationship in cases where factor Z1 is absent). If neither holds, there is the significant risk of an *ecological fallacy* when inferring from population-level trends to individual cases (Robinson, 1950). Actual experiments have the further difficulty that their inferences do not necessarily hold outside the controlled laboratory setting, meaning that the ability to infer to cases outside the lab is even further reduced. For example, in the 2003 Tomz and Weeks study discussed earlier, it is questionable whether the results gained from asking respondents in a survey about a hypothetical actually are transferable to real-world crisis situations in which we should expect a range of other factors to be at play (e.g. media and partisan priming).

However, even if we were able to estimate accurately propensity scores for individual cases, studying causal claims by comparing values of X and Y *across* cases would not tell us *how* causes work *within* a case. In other words, we learn about the difference variation in X makes for values of Y, but we do *not* learn anything about the causal arrow linking the two – it remains firmly within a black box. An experiment does not tell us *how* a treatment works – only that there is a mean causal effect (Dowe 2011; Illari 2011; Machamer 2004; Russo and Williamson 2007; Waskan 2011). Similarly, if we are interested in processes of relational meaning-making, the experiment tells us nothing about how this works.

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 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.

However, because case-based research makes case-level ontologically deterministic causal claims, knowledge about detailed causal mechanisms that are operative within single cases cannot easily be exported to other cases because the operation of causal processes are typically sensitive to even slight contextual differences (Falleti and Lynch 2009; Gerring 2010; Goertz and Mahoney 2009; Ragin 2000).

This means that, using Cartwright's language (2012), in case-based research we learn about how 'it works here', but it becomes very difficult to extrapolate our findings to other cases that it also 'works there' because of the importance of contextual conditions. This problem can be termed *mechanistic heterogeneity*, where either 1) the same cause triggers different mechanisms that are linked to different outcomes, or 2) the same cause and outcome are linked together through different mechanisms in different contexts. The risk of the first variant of mechanistic heterogeneity can be reduced through careful mapping of the population by scoring cases on their values of the cause, outcome and contextual conditions. However, the second scenario is more problematic because mechanistic heterogeneity might be lurking under what might look like a homogeneous set of cases at the level of causes/outcomes. Given this sensitivity, our ability to generalize about mechanisms from studied cases to other cases using comparisons is significantly weakened. In effect, in case-based research we trade higher internal validity of mechanistic causal inferences for a more limited ability to generalize beyond the studied population (i.e. lower external validity). Extrapolating from the individual (or small group) to the full population in this situation would result in an *atomist fallacy*.

One alternative to taking mechanistic heterogeneity seriously by appreciating the complexity of real-world cases would be to lift the level of abstraction about our theorized mechanisms to such a high level that our theorized mechanisms tell us precious little, if anything, about how a process works in real-world cases. For example, Elster talks about mechanisms such as 'wishful thinking' or 'the spillover effect' that tell us nothing about the actual causal process and how it works in real-world cases (1998: 45, 52–55). Instead of lifting the level of abstraction to the level of a one-liner, case-based scholars make more extensive claims about processes operative in smaller, bounded sets of cases (Bechtel and Richardson 2010). Appreciating complexity does not mean that we cannot engage in cumulative research. Ideally, after intensive collaborative research over a longer time period, the result would be an evidence-based catalogue of different mechanisms that are triggered by a given cause (or set of causes) in different contexts.

In theory, detailed, case-based process-tracing of mechanisms could then be combined with more contextualized variance-based controlled comparisons - for example by doing an experiment on a relatively homogenous sub-population that had also be studied using process-tracing. However, this then raises an additional problem - 'masking' (Steel, 2008: 68; Clarke et al 2014). Masking means that a given cause might be linked to the same outcome through *multiple* mechanisms that can have *different* effects on the outcome. For instance, exercise (X) triggers two different mechanisms that are related to weigh (Y): one related to weight loss through burning calories, and the other related to weight gain through building of muscles. Tracing the 'burning calories' mechanism between exercise and weight loss in a case tells us that there is a causal mechanism linking the two, but it is uninformative about the net causal effects. If we are interested in mechanisms, we are typically answering a 'how does it work' question, whereas the variance-based approach would be answering a '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 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 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. Dehumanizing social science by turning it into 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.

Secondly, 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.⁸

⁸ Fredric C. Schaffer (2014) problematizes generation and analysis of data about democracy relying on thin descriptions coming from 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).

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.

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 ph.d. 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 to least understand and appreciate the core differences across the 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 from variance- and case-based approaches (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 variance- or case-based (e.g. Bevir and Rhodes, 2010, 2016; Yanow and Schwartz-Shea, 2012, 2014).

Finally, we do not suggest that multi-methodological research is completely impossible, but it will typically require collaboration between scholars fluent in different methodologies. Therefore, we suggest that multi-methodological research be undertaken by teams of scholars who have specialized in their respective methodologies. The goal would though 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: Princeton University Press.
- Anscombe, Gertrude E M. 1957. *Intention*. Oxford: Basil Blackwell.
- Bamanyaki, Patricia A. and Nathalie Holvoet. 2016. Integrating theory-based evaluation and process tracing in the evaluation of civil society gender budget initiatives. *Evaluation* 22(1):72 - 90.
- Beach, Derek, and Rasmus Brun Pedersen. 2016. *Causal Case Studies: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor: University of Michigan Press.
- Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process-tracing methods*. 2nd edition. 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 Colin Elman. 2006. Complex Causal Relations and Case Study Methods: The Example of Path Dependence. *Political Analysis* 14(2):250–267.
- Berg-Schlosser, Dirk. 2012. *Mixed Methods in Comparative Politics: Principles and Applications*. Houndmills: Palgrave Macmillan.
- 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 R.A.W. Rhodes. 2010. *The State as Cultural Practice*. Oxford: Oxford University Press.

- Bevir, Mark, and Roderick A W Rhodes. 2016. *Routledge Handbook of Interpretive Political Science*. London: Routledge.
- Bogen, Jim. 2005. Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 397-420.
- Brulé, David J.. 2008. Congress, Presidential Approval, and U.S. Dispute Initiation. *Foreign Policy Analysis*, 4(4): 349–370.
- Bunge, Mario. 1997. Mechanism and Explanation. *Philosophy of the Social Sciences* 27(4): 410-465.
- 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: Palgrave.
- Cartwright, Nancy. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.
- Cartwright, Nancy. 2011. Predicting 'It Will Work for Us': (Way) beyond Statistics. In *Causality in the Sciences*, ed. Phyllis McKay Illari, Federica Russo, and Jon Williamson, 750–68. Oxford: Oxford University Press.
- Cartwright, Nancy. 2012. Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps. *Philosophy of Science* 79(5): 973-989.
- Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
- 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.
- Collier, David, and James Mahoney. 1996. Research Note: Insights and Pitfalls: Selection Bias in Qualitative Research. *World Politics* 49 (1): 56–91.
- Dafoe, Allan, John R. Oneal and Bruce Russett. 2013. The Democratic Peace: Weighing the Evidence and Cautious Inference. *International Studies Quarterly*, 57(1): 201-214.
- Dowe, Phil. 2011. The causal-process-model theory of mechanisms. In Phyllis McKay

- Illari, Federica Russo and Jon Williamson (eds) *Causality in the Sciences*. Oxford: Oxford University Press, 865-879.
- 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.
- Falleti, Tulia G. and Julia F. Lynch. 2009. Context and Causal Mechanisms in Political Analysis. *Comparative Political Studies* 42:1143-1166.
- 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*, eds. Walter Carlsnaes, Thomas Risse and Beth A. Simmons. London: SAGE: 52-72.
- Fujii, Lee Ann. 2008. 'The Power of Local Ties: Popular Participation in the Rwandan Genocide.' *Security Studies* 17(3): 568–97.
- Geertz, Clifford. 1973. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.
- Geertz, Clifford. 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 Bennett Andrew. 2005 *Case studies and Theory Development in the Social Sciences*. Cambridge, Massachusetts: MIT Press.
- Gerring, John. 2010. Causal Mechanisms: Yes But... *Comparative Political Studies* 43(11): 1499-1526.
- Gerring, John. 2011. *Social Science Methodology—a unified framework*. Cambridge: Cambridge University Press.
- Gerring, John. 2017. *Case Study Research*. 2nd edition. Cambridge: Cambridge University Press.
- Goertz, Gary. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton: 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 by David Byrne and Charles C. Ragin (eds) *The Sage Handbook of Case-Based Methods*, Thousand Oaks: SAGE, pp. 307–17.
- Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press.
- Groff, Ruth. 2011. Getting past Hume in the philosophy of social science. In Phyllis McKay Illari, Federica Russo and Jon Williamson (eds) *Causality in the Sciences*. Oxford: Oxford University Press, 296-316.
- Hansen, Lene. 2006. *Security as Practice: Discourse Analysis and the Bosnian War*. New York: Routledge.
- Haggard, Stephan, and Robert R. Kaufman. 2016. *Dictators and Democrats: Masses, Elites, and Regime Change*. Princeton: Princeton University Press.
- Holland, Paul W. 1986. Statistics and Causal Inference. *Journal of the American Statistical Association* 81(396): 945-960.
- Hopf, Ted. 1998. 'The Promise of Constructivism in International Relations Theory.' *International Security* 23(1): 171-200.
- Hopf, Ted. 2002. *Social Construction of International Politics: Identities & Foreign Policies, Moscow, 1955 and 1999*. Ithaca: Cornell University Press.
- Humphreys, Macartan, and Alan Jacobs. 2015. Mixing methods: A Bayesian approach. *American Political Science Review* 109(04):653–673.
- 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*, ed. 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.
- Jackson, Patrick Thaddeus. 2014. 'Making Sense of Making Sense: Configurational

- Analysis and the Double Hermeneutic.’ In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*, eds. Dvora Yanow and Peregrine Schwartz-Shea. New York: M.E. Sharpe, 267–83.
- Jackson, Patrick Thaddeus. 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, Hoyoon. 2019. ‘The Evolution of Social Constructivism in Political Science: Past to Present.’ *SAGE Open*, January-March: 1-10.
- 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*, DOI: 10.1080/1350178X.2018.1561078
- King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Kreuzer, Markus. Forthcoming. *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 2000-2001. Contingency, Catalysts and International System Change. *Political Science Quarterly* 115 (Winter 2000–2001): 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*. 2006;7(Suppl 1):S1. doi: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, DOI: 10.1080/09636412.2015.1070602

- 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.
- Lijphart, Arend. 1971. Comparative politics and the comparative method. *American Political Science Review* 65 (3): 682–93.
- 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): 147-184.
- Pedersen, Jonas Gejl. 2019. *The Russian Quest for Ontological Security*. Aarhus: 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, Los Angeles, London: University of California Press.

- Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Robinson WS. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357.
- Rubin, Donald. 1980. Discussion of ‘Randomization analysis of experimental data in the Fisher randomization test’. *Journal of the American Statistical Association*, 75 (371): 591-593.
- Rubin, Donald 2005. Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 100(469): 322-331.
- Runhardt, Rosa W. 2015. Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward’s Manipulability Theory of Causation. *Philosophy of Science* 82 (5): 1296-1307.
- Russo, Federica, and Jon Williamson. 2007. Interpreting Causality in the Health Science. *International Studies in the Philosophy of Science*, 21 (2): 157–70.
- Russo, Federica, and Jon Williamson. 2011. Generic versus Single-Case Causality: The Case of Autopsy. *European Journal of the Philosophy of Science* 1 (1): 47–69.
- Schaffer, Frederic Charles. 2014. ‘Thin Descriptions: The Limits of Survey Research on the Meaning of Democracy.’ *Polity* 46(3): 303–30.
- Schaffer, Frederic Charles. 2016. *Elucidating Social Science Concepts: An Interpretivist Guide*. New York: Routledge.
- Schmitt, Johannes and Derek Beach. 2015. The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, 21(4): 429–447.
- 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.
- Schneider, Carsten Q., and Ingo Rohlfing. 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 *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*, New York: M.E. Sharpe, 120–46.

- Seawright, Jason. 2016. *Multi-Method Social Science*. Cambridge: Cambridge University Press.
- Steel, Daniel. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University 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: Princeton University Press.
- Tomz, Michael R. and Jessica L. P. Weeks. 2013. Public Opinion and the Democratic Peace. *American Political Science Review*, 107(4): 849–865.
- Waskan, Jonathan. 2011. Mechanistic Explanation at the Limit. *Synthese* 183 (3): 389–408.
- Walters, Benedict and Derek Beach. 2018. Process tracing and congruence analysis to support theory based impact evaluation. *Evaluation*, 24(3): 284–305.
- Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- Williams, Malcolm and Wendy Dyer. 2009. Single case probabilities. In: Ragin and Byrne (eds) *Case Based Methods*. London: SAGE, pp. 84–100.
- 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.
- Yanow, Dvora, and Peregrine Schwartz-Shea. 2014. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. 2nd ed. eds. Dvora Yanow and Peregrine Schwartz-Shea. New York: M.E. Sharpe.