Welcome Letter by QMMR Section President Melani Cammett

I am delighted to serve as the President of the QMMR section, which plays a key role in the discipline as a meeting point for scholars with diverse intellectual and methodological orientations. The vibrancy of the section attests to the fact that qualitative and multi-methods research continue to thrive in political science.

The section officers deserve special credit for carrying out the work of the section. I especially want to acknowledge Peter Hall, who recently completed his term as QMMR President. Among his many contributions, Peter actively promoted the deliberations around research transparency and integrity—an activity that has been a critical focus of the section over the past couple of years. As ever, Colin Elman, the QMMR Secretary-Treasurer, works tirelessly on behalf of the section. It would not be an exaggeration to say that, without him, we would not be where we are today.

Special thanks are also due to Tim Büthe and Alan Jacobs for their important work on two fronts—as editors of Qualitative and Multi-Methods Research and for spearheading the Qualitative Transparency Deliberations (QTD). First, during their tenure as co-editors, Tim and Alan continued to elevate the profile of our section’s flagship publication by publishing a variety of symposia on important topics. Thanks to their efforts, building on the work of their predecessors as editors of QMMR, it has become a scholarly publication rather than just a “newsletter,” a label which the section appropriately decided to drop from the title in 2016. The recent adoption of DOIs for articles published in the newsletter symbolizes the very real intellectual contributions of the symposia and articles featured in this publication. In addition, at the QMMR annual business meeting at the APSA meetings held in 2017, the membership voted to switch the newsletter to an all-digital format under the incoming co-editors, Jennifer Cyr and Kendra Koivu. Among other benefits, this will allow for more funds to be devoted towards editorial work for this valuable resource.

continued on page 58
Truth Seeking AND Sense Making: Towards Configurational Designs of Qualitative Methods

Joachim Blatter
University of Lucerne, Switzerland

Introduction and Overview

Within the qualitative social sciences, we can detect a wide gulf between those who strive for revealing "the truth" about the social world on the one hand and those whose goal is to "make sense" of it on the other. The former apply methods which are implicitly or explicitly rooted in positivist or realist epistemologies and ontologies whereas the latter apply methods based on constructivist or interpretative epistemologies and ontologies. Some of the characteristic expressions of this gulf can be found in the work of Goertz and Mahoney who exclude interpretative approaches in their characterization of qualitative research. This is mirrored by Yanow who insists on the distinctiveness of interpretative research.

Nevertheless, a closer view reveals that on both sides, among ‘truth-seekers’ and among ‘sense-makers,’ we find quite distinct research goals, epistemological principles, and ontological presumptions, as well as a broad spectrum of methods of data collection/creation and methods of data analysis/interpretation. The major claim of the following contribution is that the internal diversity within both camps makes it possible to develop a plurality of coherent qualitative methods which allow to strive for truth-seeking and sense-making at the same time. These methods are configurational in the sense that they combine epistemological and ontological features of truth-seeking endeavors with those of sense-making projects. If appropriately conceptualized and designed, they do this without losing their internal coherence and are therefore helpful for building bridges across the aforementioned gulf.

In order to develop these configurative and coherent methods, I start by reflecting on epistemology and argue that a pragmatic epistemological stance consists of three components: a) a research goal specified by a research question; b) the corresponding kind of aspired knowledge/explanation; and c) the adequate way to secure the validity (and reliability) of the research process and its results. For each component, I identify principled differences between truth-seekers and sense-makers, but point to internal diversity on each side, as well. Next, I turn towards ontology and sketch the three components of a pragmatic ontological stance: a) presumptions about the basic entities of the social world; b) presumptions about the relationship among these entities; and c) understandings of causes and causation. Once again, I identify principled differences between the two camps and internal diversity in each camp. Based on these premises, I present two tables in which the scrutinized alternatives are aligned in such a way that six ideal typical methods emerge. Each of these methods is characterized by an internally coherent set of epistemological purposes/principles and ontological presuppositions. A third table in which I lay out the corresponding methodological approaches to data collection/creation and to data analysis/interpretation can be found elsewhere. Due to the space limitations of this paper, it could not be incorporated.

Figure 1 provides a first overview, indicating not only the conceptual poles in the epistemological and ontological dimensions, but also the location of the six ideal typical methods. Furthermore, figure 1 specifies and substantiates the major claim of the contribution formulated above. I argue that an ideal typical method called Comparable Cases Strategy strives single-mindedly for revealing the truth (about the autonomous effect of a single cause). In contrast, the Con-Textual Analysis method aims entirely at making sense by creating enlightening understandings of the empirical world. The four methods in between, though, can be conceptualized in such a way that they combine truth seeking and sense making without being incoherent. Methodological approaches labelled Configurational Comparative Analysis and Co-Writing Cultures are still predominantly committed to either truth seeking (the former) or to sense making (the latter), but they incorporate already some epistemological and ontological features of the other side. Causal-Process Tracing and Congruence Analysis (two specified forms of within-case analysis which are often—alongside other forms of within-case analysis—lumped together under the term ‘Process Tracing’), represent methods that come close to balancing the epistemological principles
Figure 1: Locating Ideal-Typical Qualitative Methods
According to Their Coherent Positions in Respect to Epistemology and Ontology

Source: Blatter, Haverland and von Hulst 2016: xxi

and ontological presuppositions of truth-seekers and sense-makers.

Figure 1 reveals that my typology of methods contains only ideal-types and not all logically possible combinations. Ideal-types are characterized by a conceptually coherent combination of specific epistemological purposes and principles on one side and specific ontological presumptions on the other. This means that I do not claim that these types exist in their ideal-typical form as methodological devices in textbooks or as distinct cultures among practitioners. At the end of this essay, I point to potential uses of a typology of ideal-types.

Before I start, I would like to signal to the reader that the attempt to bridge a very broad spectrum of qualitative methodologies and to develop ideal-types has two consequences in respect to terminology:

1. I use core expressions like “explanation” and “causation” as umbrella terms. This means that I treat “interpretation” and “understanding” as specific types of explanation and “constitution” or “construction” as a specific type of causation. Such a stance can be justified as follows: A common terminology facilitates mutual understanding and helps to build bridges between truth-seeker and sense-makers. Furthermore, terms and terminology frame the (scientific) discourse and set the limits of what is accepted and what is not. Thus, if interpretation is not accepted as a form of explanation, or if causality is limited to the empiricist/analytic understandings that dominate the natural sciences, then these are fundamental forms of exclusion.

2. The terms that are used to label the ideal-typical methods are based on a broad reading of the literature and are picked in order to correspond as much as possible to existing terminology. Nevertheless, the ideal-typical approach implies that the main criteria for choosing labels has not been correspondence to existing terminology, but linguistic coherence and conceptual correspondence to the scrutinized features of the ideal-type.

Towards Pragmatic and Pluralist Epistemological Stances

In his widely cited textbook Approaches to Social Enquiry, Blaikie defines epistemology as “a view and justification for what can be regarded as knowledge—what can be known, and what criteria such knowledge must satisfy in order to be called knowledge rather than belief.” Such a definition tends to lead to principled disputes about what we can know in the social sciences given the facts that social scientists are parts of the social world they study, to the extent that some scholars claim that the social word cannot be studied as an externally or objective existing entity. A pragmatic approach, in contrast, starts with the assumption that we should accept all kinds of research goals which can be specified in distinct research questions. For each research question, we can identify an understanding of knowledge that is most appropriate for providing a useful answer to the question, and we can specify the criteria that guide the processes of gaining this kind of knowledge. As

---

5 In contrast to Wendt 1999, but in line with Elster 2007 and Kurki 2008.

6 Blaikie 1993, 7.
we will see, the most important advantages of such a pragmatic approach are that it helps to overcome dichotomous thinking and that it paves the way towards epistemological pluralism.

If we accept that an epistemological stance should be based on what we want to know (instead of what we can know given some specific ontological presupposition, for example), then a specific epistemological stance encompasses three components:

a. the research goal expressed in a precise research question and translated into a corresponding research design,

b. the type of knowledge/explanation that we aspire in order to answer the research question (e.g. if we want to know, how - and not whether or how much - X influences Y, then we strive for a mechanism-based explanation), and

c. the principles and procedures that guide the process of acquiring this kind of knowledge and the corresponding criteria for evaluating the quality of concrete research projects.

In the following section, I do not only scrutinize the principled alternatives in respect to these three components, but I indicate how we can overcome the dualism that is invoked by focusing just on the principled alternatives.

Research Goals: Truth-Seeking, Sense-Making and their Combinations

Social scientists, especially those who pretend to do “qualitative” research, find themselves located between the hard/natural sciences and the arts/humanities. Therefore, it is not surprising that some qualitative social scientists—in line with natural scientists—adhere to “truth-seeking” as their principled goal of research, whereas others strive—in line with those from the humanities—for “sense-making.” The prototypical research goal for truth-seekers is to develop and test parsimonious hypotheses and models that correspond to the main features of an external world. In contrast, the prototypical research goal of the sense makers is to develop and apply coherent paradigms and theories that provide orientation through meaningful interpretations of the world. In consequence, it seems that truth-seekers and sense-makers have clearly distinct and seemingly incommeasurable research goals. Nevertheless, each of these principled research goals allows for a range of possible specifications as to what a research question is.

Among the prototypical research questions that truth-seekers might try to answer are:

a) Which effects does a specific and concrete cause (X) have?
b) Which configurations of conditions make a specific kind of outcome (Y) possible?
c) Which underlying mechanisms (M) make causes produce an effect?

Sense-makers ask questions that also take a wide range of prototypical forms, such as:

A) Which fundamental structure (S) stabilizes and/or transforms a social/political system?
B) Which interpretative signs and practices characterize a specific culture (C)?
C) Which paradigms-based, but specified theory (T) provides a better explanation?

These prototypical research questions signal the existence of a plurality of specified research questions within each camp. In the following section, I want to show that the search for answers to some of these questions demands research designs and methods that combine truth seeking and sense making. For example, I argue that the most productive answer to question c) is based on an understanding of a causal mechanism as a configuration of three kinds of social mechanisms. Furthermore, it demands a kind of Causal-Process Tracing that aims to show not only that the identified mechanisms correspond to an external reality, but also that they make sense insofar as the social mechanisms are integrated in a coherent and comprehensive multi-level model of explanation. Likewise, to answer question C), we need a method that strongly combines sense-making and truth-seeking. As we will show later on, the corresponding method, Congruence Analysis, draws on the abstract approach to theory-formation that sense-makers adhere to, but the answer also depends on the systematic comparison of the expectations that we can deduct from those theories with the kinds of empirical observations that truth-seekers demand in order to accept an explanation.

At this point, these statements are not much more than claims, but I try to clarify and justify them in the following section. The first step on this path is to show that we find a similar plurality (instead of a dichotomy) when it comes to the principled kinds of knowledge that we are striving for and when we develop criteria for gauging the quality of the process through which we gain this kind of knowledge. Because most methods in the social sciences have been developed as tools of explanation, we limit our discussion to explanatory knowledge, although we generally share John Gerring’s view that “descriptions” and “comparisons” constitute forms of knowledge that are at least as important.

Explanations: Confirmed Thesis, Convincing Paradigm, and Options In-between

In the social sciences, there are multiple and quite different understandings of what a ‘good’ explanation is. In order to overcome simple dichotomies without erasing fundamental differences, we will develop a two-dimensional space for locating distinct understandings of explanations. The first dimension of this conceptual space refers to the level of abstractness and the second dimension to the level of generality. Since Collier and Mahon’s path-breaking work on concept building.

7 Blatter, Langer and Wagemann 2017, 7-17.
8 Gerring 2012.
9 Collier and Mahon 1993.
we know that Sartori’s “ladder of abstraction” was a misnomer: abstraction and generalization are not the same thing, and Sartori was primarily concerned with the problems of generalization. The opposite of “abstraction” is “concretization”, whereas the opposite of “generalization” is “specification.”

The dichotomy between abstract and concrete concepts shows up in the distinct procedures through which these two kinds of concepts are defined: Abstract concepts are defined through reflection on the relationships that one concept has to other abstract concepts. The attributes that we select for characterizing our abstract concept have to be justified with reference to a theoretical discourse. Concrete concepts, in contrast, are defined through the assignment of indicators that refer to observations. The categorical difference between abstract and concrete concepts shows up, once again, when we look at how the “negative pole” of the concept is getting defined. For an abstract concept, the negative pole must be defined through a substantial alternative concept (e.g. “monarchy” or “autocracy” for the concept of democracy). For a concrete concept, though, it is enough to define the negative pole as simple negation or as the null point (non-democracies, zero degree of democracy).

When we reflect about the level of generality of a concept, we are not concerned with how the concept’s defining characteristics have been derived. Instead, we reflect on the relationship between the set of characteristics or attributes that define a concept (“category” in Collier and Mahon’s terminology) and the set of entities in the world to which the concept refers. The former is called the “intension” of a concept, the latter the “extension.” Collier and Mahon’s most important insight is that only in classical systems of categorization does a higher intension (a concept that is more specified by a higher number of attributes) lead to a lower extension (a lower number of entities that correspond to the concept). If we use family resemblance or radial categories (systems of categorization in which some attributes are possible but not necessary attributes of a category), there is no logical trade-off between intension and extension. This means that the extension of a category that is located on a lower level of generality may exceed that of a category that is located on a higher level of generality. This is because, in these systems of categorization, going down the ladder of generality means to select a specific configuration out of a larger set of possible attributes which characterize a concept.

In line with such non-classical systems of categorization, we can define

a) a Paradigm (P) as the set of all theories that combine a specific core concept (CC) with one or a plurality of peripheral concepts (PC): \( P = CC \ast [PC1 + PC2 + PC3 + \ldots] \),

b) a Theory (T) as a specific combination of the core concept with one or a plurality of peripheral concepts: \( T = CC \ast PC1 \ast PC2 \).

The formulas reveal that a Paradigm has not only a larger intension (more attributes) but also almost certainly a larger extension (more empirical entities to which the concepts refers) than a Theory. In consequence, we might redefine what “intension” means for non-classic systems of categorization and for non-classic approaches to concept and theory formation: In these contexts, “intension” would not refer to the number of attributes that characterize a concept, but to the intensity by which these attributes are linked to each other. If we accept this definition of intension, we end up also with what intuitively makes sense for the non-classic systems of categorization: a higher level of intensity leads to a lower level of extensity. The meaning of this insight, however, is markedly different to the currently dominant understanding which is still in line with the writings of Sartori.

So far, we have clarified the difference between abstraction and generalization. Furthermore, we have pointed to the categorical differences between abstract and concrete concepts, and introduced non-classic forms of categorization or concept formation. In earlier contributions, we transferred these insights from concept-formation to the task of theory-formation. Whereas “attributes” represent the elements that we use for specifying “concepts,” “concepts” are the elements that we use to specify “theories.” Such a transfer paves the way to conceptualizing the relationship between paradigms and theories in the language of concept formation.

In the following, we build on these insights and turn towards a systematic mapping of the different types of explanation for which the different strands of research strive to (see figure 2).

On the one hand, we might want to find out whether a concrete cause (low level of abstraction) has a specific effect (low level of generality). The corresponding hypothesis that provides a preliminary answer can be deduced from abstract theories, but it does not have to. Quite often, such a hypothesis is (seen as) nothing more than an unproven claim. In order to test the causal claim of said hypothesis, the independent and the dependent variable must be clearly specified and operationalized (concretized) by observable indicators. If we can control for all other potential causal factors by comparing similar cases in these respects, the observed co-variation among the independent and dependent variable provides enough empirical leverage for transferring the hypothesis into a truthful thesis about a causal relationship within a clearly delimited population of cases.

On the other hand, there are abstract paradigms, which aspire to make sense and provide orientation for many instances
Figure 2: Different Kinds of Explanation

of social entities and for many facets of the social world (for paradigms, there are no boundaries of the population of cases to which they refer to). They are characterized by core concepts and a large set of peripheral concepts, whereby both the core concepts and the peripheral concepts remain on an abstract level, so that it needs a lot of interpretative work in order to connect empirical observations to these abstract concepts.

For an understanding of the difference between the other two types of explanations (theories, models), it is helpful to perceive them as less radical siblings of theses and paradigms. Like (hypo-)theses, models are located on a concrete level, but they are not as narrowly specified—they take a broader set of causal factors into account for explaining the outcome. This stands in contrast to when we want to test a hypothesis. In such cases, we try to control for most factors and to focus on one single independent variable and one dependent variable. A model can be a statistical model (Dependent Variable $Y = a \cdot \text{Independent Variable X1} + b \cdot \text{Indep. Variable X2} + c \cdot \text{Indep. Variable X3} + \text{error}$), a set-theoretical model (Outcome = Condition A * Con. B + Con. C * Con. D), a causal chain (Precondition A -> Precond. B -> Precond. C -> Outcome) or a multi-level model of a causal mechanism (Causal Mechanism = Situational Mechanism * Action-formation Mechanism * Transformational Mechanism). Crucially important—and the main difference to a theory—is the fact that models are integrated on an empirical level. A good model has a good “fit” to the empirical data. The various elements of the model do not have to be conceptually consistent in the sense of belonging to a single worldview.

Such a conceptual coherence is exactly what characterizes a theory in contrast to a model. As we have laid out before, a theory is a specified paradigm in the sense that it combines one or a few peripheral concepts and the core concept of the paradigm and specifies the status of the selected peripheral concepts as necessary conditions for the theory ($\text{Theory} = \text{Core Concept} \ast \text{Peripheral Concept a} \ast \text{Peripheral Concept b}$). This means that a theory is located on a lower level of generality, but it remains abstract in the sense that its conceptual elements are derived first and foremost by discussing their relationship to other abstract concepts and not by referring to (existing) indicators. Crucially important for an adequate understanding of the kind of explanation that a theory provides is that the conceptual elements of a theory form a coherent whole through their belonging to a common worldview/paradigm. Elements are only included into an explanatory framework if they conceptually fit with the other elements of the theory and not if they enhance the fit to the empirical data.

Overall, the differentiation between levels of abstraction and levels of generality allows us to develop a pluralistic view on distinct kinds of explanations. The former aspect reflects the differences-in-kind between truth-seekers who strive for explanations on a low level of abstraction and sense-makers who prefer explanations of a higher level of abstraction. The latter aspect makes us aware that there are differences-in-degree on both sides: Truth-seekers develop and not only test hypotheses which focus (ideal-typically) on causal relationships between a single independent variable of interest and the dependent variable in a very limited population of cases, but also models which include causal relationships among a plurality of variables, conditions or mechanisms in a larger population of cases. Sense-makers, in turn, do not only develop and apply paradigmatic lenses that provide insights and orientation in general, but also theories that are more tailored for specific contexts.
Can we trust the results?

Usually, truth-seekers and sense-makers interpret and specify these criteria quite differently: In respect to validity, the former argue that we have to make sure that we describe and explain what we want to describe/explain by relying on formal logic. Principles of formal logic are “objective” and independent from the standpoint of the applicant. In consequence, and in line with the goal to seek the truth, the first approach to validity aims at “objectivity” and prescribes “neutrality” for the researcher. Sense-makers, in contrast, argue that we have to secure adequate descriptions and explanations with the help of the associative and justificatory faculties that languages offer. These faculties depend on, and are shaped by, the specific language (theoretical lenses, concepts) that the researcher applies. In consequence, good research has to justify the selected theories and concepts by reflecting on their position in the scientific discourse and in the social/political practice and their relationship to other theories and concepts.

Nevertheless, we get a more nuanced picture of the meanings of validity, and of how validity is sought, if we break down the analytic process into four components that are necessary for producing an explanation. For a comprehensive, line with truth-seeking, the second strategy with sense-making.

I briefly scrutinize the different strategies at the four focal points along the following sequence of a deductive research process: a) concept specification; b) concept concretization (operationalization); c) conclusions from the created data to the relationships among the concepts for the studied cases; d) reflections on the wider generalizations of these findings.

The validity of the specification of a concept depends on how (much) we justify the assignment of specific attributes by relating the selected concept to other concepts within the scientific discourse. Sometimes this procedure is presented as involving two steps: First, a “systematized concept” is derived from a “background concept:” the selection of a specific meaning from the universe of possible meanings is justified with the specific goals or purposes of the research project. Second, indicators are selected which “represent the universe of content entailed in the systematized concept.”

Truth-seekers adhere to such a content-centered approach to concept formation since it allows them to treat concepts as clearly externally-delineated and internally homogeneous elements. Sense-makers, in contrast, emphasize the context-dependent meaning of concepts and the intersubjective construction of meaningful concepts. Accordingly, for them, the specification of a concept involves a reflection on the position and the role of a concept (its linguistic signifier) in the scientific discourse. The internal characteristics of the concept are not determined by selecting the best observable representative of a homoge-

---

**Table 1: Different Focal Points and Different Strategies for Securing Validity**

<table>
<thead>
<tr>
<th>Securing validity for the cases under study</th>
<th>Description</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>b) Valid Concretization: Selecting <strong>convergent OR alternative</strong> indicators</td>
<td>c) Valid Conclusion: Linking abstract relationships to concrete observations through <strong>inference OR interpretation</strong></td>
<td></td>
</tr>
<tr>
<td>a) Valid Specification: Justifying attributes by referring to <strong>content OR context</strong></td>
<td>d) Valid Generalization: Presuming <strong>causal OR constitutive</strong> scope conditions</td>
<td></td>
</tr>
</tbody>
</table>

**Quality Criteria: Different Ways to Secure Validity Between Neutrality and Positionality**

Validity and reliability are the most basic quality criteria for research procedures. These two criteria point us to the most important questions for judging our process of knowledge creation: Do we study/explain what we claim to study/explain? Can we trust the results?

For a comprehensive, line with truth-seeking, the second strategy with sense-making.

When we combine the two dimensions, we get four focal points that indicate where we try to secure validity (see Table 1). Furthermore, Table 1 highlights that for each focal point we can detect two distinct strategies for pursuing the corresponding task of validation. Within each cell, the first strategy is in line with truth-seeking, the second strategy with sense-making.

I briefly scrutinize the different strategies at the four focal points along the following sequence of a deductive research process: a) concept specification; b) concept concretization (operationalization); c) conclusions from the created data to the relationships among the concepts for the studied cases; d) reflections on the wider generalizations of these findings.

The validity of the specification of a concept depends on how (much) we justify the assignment of specific attributes by relating the selected concept to other concepts within the scientific discourse. Sometimes this procedure is presented as involving two steps: First, a “systematized concept” is derived from a “background concept:” the selection of a specific meaning from the universe of possible meanings is justified with the specific goals or purposes of the research project. Second, indicators are selected which “represent the universe of content entailed in the systematized concept.”

Truth-seekers adhere to such a content-centered approach to concept formation since it allows them to treat concepts as clearly externally-delineated and internally homogeneous elements. Sense-makers, in contrast, emphasize the context-dependent meaning of concepts and the intersubjective construction of meaningful concepts. Accordingly, for them, the specification of a concept involves a reflection on the position and the role of a concept (its linguistic signifier) in the scientific discourse. The internal characteristics of the concept are not determined by selecting the best observable representative of a homoge-
neous concept but by reflecting on the (categorical and consequent) relationships of the concept to other abstract concepts. 19

The validity of the **concretization** (often called operationalization) of a concept depends primarily on the selection of the correct indicators. The most important test that confirms this is whether the scores produced by an indicator are empirically associated with scores of other (direct) measures of the concept. This kind of validity is often called “criterion validity” and the procedure is labeled “convergent validation.” 20 Nevertheless, Gary Goertz has made us aware of the fact that the scores of indicators should converge only if we believe the indicators to be consequences of our concept (which would be termed a latent variable in quantitative research). 21 If we perceive the relationship between indicator and concept not as causal (in a narrow sense, see below) but as functional or constitutive (as sense makers often do), different valid indicators of a concept need not converge because they may be understood as alternative options for making the concept possible.

The validity of the explanations that we derive for the cases we study depends primarily on whether the conclusions that we draw from observations/signs to unobservable relationships between our concepts are consistent from the viewpoint of formal logic or whether they are coherent in the sense that they are convincingly justified (explicitly, with means with the help of language). The former is denoted by the term **inference** (truth-seeking), the latter by the term **interpretation** (sense-making).

The validity of the **generalizing** conclusions that we draw from our results depends on the adequacy of some fundamental presumptions. Once again, we can detect different procedures for strengthening what is also called external validity. **Construct validation** refers to procedures which start with the presumption that specific causal relationships exist. 22 For example, the Comparable Cases Strategy depends strongly on presumptions about other factors (beyond the factor of interest) that might influence the dependent variable. Already the validity of the conclusions for the cases under study crucially depends on the correct identification of alternative variables of influence. The only way to control for the influence of these factors is to take them as criteria for case selection (the selected cases must show no variation in respect to these factors). The same is true when it comes to draw generalizing conclusions beyond the studied cases. We can generalize the result of our cross-case analysis only for the population of cases that show similar values in respect to the control variables as our selected and analyzed cases, because only within this—often very small population of cases which share the same scope conditions—we can be sure that our factor of interest is responsible for a causal effect and not another factor.

In order to highlight the functional equivalency, and in line with our valuation of linguistic coherence, we might call the principled alternative to construct validity “**construction validity.**” Like the former, the latter depends on a presumption of relationships between the scope conditions and our outcome of interest. Whereas the former presumes causal relationships in a narrow sense, the latter presupposes that the specific material or ideational structures which are being focused on have a constitutive effect on social actions and processes. For example, those who analyze discourses or narratives presume that these linguistic structures have a constitutive impact on the interests and interactions of social/political actors. In a Con-Textual Analysis, the analytic focus is not on testing this presumption but on the formation and transformation of these structures. In combination with the objectivists versus conventionalist understandings of knowledge, to which truth seekers and sense makers respectively adhere to, we can conclude: generalizations of truth-seekers depend on the truth of their causal presumptions in respect to control variables/ scope conditions whereas generalizations of sense-makers, in contrast, depend on how well the constitutive presumptions are accepted in the scientific discourse.

**Towards Pragmatic and Pluralist Ontological Stances**

According to Blaikie, ontology refers to “the nature of social reality—claims about what exists, what it looks like, what units make it up and how these units interact with each other.” 23 Blaikie’s definition draws him immediately into the philosophical debate on whether the social reality that we study exists independent of the human mind. Quite similarly, Brady associates ontology primarily with the question of deterministic versus probabilistic causality. 24 A **pragmatic** stance, in contrast, starts by emphasizing that it is legitimate to ask all kinds of research questions, despite the fact that these questions imply very different assumptions about the nature of social reality. For each research question, we should use those assumptions about the nature of social reality that allow us to develop the most useful answer to the question. Such a pragmatic view of ontology not only helps us to build bridges between seemingly incommensurable ways of conceptualizing the nature of social reality, but also allows us to differentiate distinct understandings of causality among truth-seekers and among sense-makers.

As we did for epistemology, we distinguish **three aspects** that are necessary in order to specify a pragmatic ontological stance: a) presumptions about the basic entities of the social world; b) presumptions about the relationship among these entities; and c) understandings of causes and causation. For each aspect, I briefly indicate the principled alternatives, but the main goal is to point to the existing plurality within each principled approach and to highlight the fact that some understandings of causality allow for a combination of truth seeking and sense making.

---

19 Adcock 2005.
23 Blaikie 1993, 6.
There are two principled alternatives when it comes to conceptualizing the basic entities of social reality: materialism and idealism. Materialists assume that the most fundamental fact about society is the nature and organization of material forces, therefore focusing on traits like the biological nature of humans, natural resources, geography, and forces of production and destruction. Idealists, in contrast, assign this role to the nature and structure of human and social consciousness, therefore concentrating on aspects like dominant forms of knowledge, ideas and values on the individual or collective level.25

For sense-makers, it is important to locate an explanatory endeavor within the general discourse about these basic entities of social reality. This is because sense-makers’ main goal is to provide orientation. Therefore, it is important to locate individual explanatory efforts within broader discourses. Doing so helps to avoid fully idiosyncratic explanations and to communicate research findings across different fields and (sub)disciplines. For truth-seekers, an explicit reflection on how the concepts they apply relate to the basic entities of social reality is not as important, since they concentrate on gauging the correspondence of explanatory models to a clearly delimited part of social reality. In consequence, truth-seekers are often agnostic with respect to the question of whether the concepts they apply in their explanatory endeavors arise from a materialist or an idealist nature of social reality. Arguably, this is the case with the concept of “actor preferences.” It is a crucial explanatory factor for rationalist explanations, but it can reflect materialist and/or idealist motivations.

Basic Entities of Social Reality: Materialism, Idealism and Beyond Binary Concepts

Causation: Different Ways to Define Causes and the Adequate Methods to Prove their Empirical Relevance

Truth-seekers usually adhere to an elementaristic understanding of causes as individual “difference makers.” A difference-making understanding of causality stipulates that causes hold a general property of making some sort of difference to their effects. On the contrary, sense-makers are more inclined to follow those who stipulate that causality is a relational concept (and not a property that a factor generally inhibits), and that the dispositional influence of a cause on the effect manifests itself only under concrete circumstances.29 In the following, I will show that when we leave the definitional level and look at the methods that are applied in order to prove the relevance of causes, we find a more diverse spectrum on both sides than such a dichotomous categorization implies.

When truth-seekers try to conceptualize causation and to identify the effects of causes, they embrace the “experimental template” as the gold standard. As Brady has shown,30 this is the case because the experimental template combines a specific understanding with an efficient way to trace this kind of causation: An experiment is based on the “counterfactual understanding” of causation expressed by Hume as “if the first object had not been, the second had never existed,”31 and allows to control two important aspects: a) the “treatment/intervention” which is necessary in order clarify the direction of the causal relation, and b) alternative factors of influence which are necessary in order to isolate the causal effect of the factor of interest.

Brady identifies two further ways to understand and to trace causation that are less focused on the identification of the effects of one specific cause. The “regularity approach” is linked to Hume’s other definition of “a cause to be an object, followed by another, and where all the objects similar to the first, are followed by the objects similar to the second.”32 It focuses on the identification of the multiple causes of a specific effect. Finally, the “mechanism approach” to causation is concerned with temporal processes and social mechanisms that link cause and effect on a lower level of analysis.33

The argument that these understandings of causation and their corresponding qualitative methods imply elementarist and relationalist ontologies can be formulated most clearly with the help of the terminology of necessity and sufficiency. When

26 Esfeld 2003.
28 Emirbayer 1997.
29 Baumgartner 2015; Anjum and Mumford 2010; Mumford and Anjum 2011.
30 Brady 2008.
31 Hume 1748 according to Brady 2008, 233.
32 Hume 1748 according to Brady 2008, 233.
33 Brady 2008, 242-245.
the counterfactual approach is applied within the **Comparable Cases Strategy (CCS)**, where we draw causal inferences from an observed co-variation of independent and dependent variables in otherwise similar cases, we presume that the cause is a necessary AND sufficient condition for the effect. If we allow the cause to be sufficient but not necessary—if we find an effect without a cause—we cannot draw any logical conclusion. Similarly, if we allow the cause to be necessary but not sufficient—if we find the cause without the effect—we cannot draw any logical conclusion either.

The regularity approach broadens this understanding by accepting conditions that are necessary but not sufficient and conditions that are sufficient but not necessary. This implies that individual causes are most often INUS-conditions—in-sufficient but necessary parts of a compound condition that is itself unnecessary but sufficient for an effect. INUS conditions imply that explanations do not contain single, “independent” factors that have “autonomous” causal power. Instead, causation involves a configuration of causal factors and that the causal power of individual factors is contingent on the existence or a specific expression of other causal factors. In consequence, the explanations that we strive for with a **Configurational Comparative Analysis (CCA)**, the method that is based on the ontological assumption of configurational causality, imply a relationalist ontology in contrast to the elementalist ontology that we presume when applying methods based on the experimental template. Nevertheless, CCA represents only a very limited step towards a holistic or relationalist ontology and its corresponding understanding of causality, since this method focuses on the empirical identification of the co-existence of a configuration of causal conditions and an outcome. It cannot provide empirical information on the mechanisms which allow the divergent components of a causal configuration to work together in such a way that they are able to produce the outcome. For that we need a different method that corresponds to an even stronger relationalist ontology.

That is exactly what we get, if we follow those methodologists who define and conceptualize the notion of “causal mechanisms” in line with major social theorists. According to such an understanding, a causal mechanism consists of three complementary social mechanisms that connect causal factors on different levels of analysis: “situational mechanisms” which link structures to actors; “action-formation mechanisms” as the micro-foundations of mechanism-based explanations; and “transformational mechanisms” which link actions, including communicative acts, to social structures. The corresponding ideal-typical method, **Causal-Process Tracing (CPT)** is designed to help find the truth in the sense of developing an explanatory model that corresponds to the empirical reality, but it is also committed to providing meaningful explanations.

This is the case not only because of the strongly relationalist understandings of causation that comes with the scrutinized conceptualization of causal mechanisms, but also because it guides the researcher towards applying one of the generic “action-formation mechanisms” that social theorists have developed. This, in turn, affirms that explanations based on Causal-Process Tracing draw on basic social theory instead of creating a flurry of idiosyncratic explanations and mechanisms, as it is the case with other approaches to Process Tracing. A final reason for assigning Causal-Process Tracing and its corresponding understanding of causation a centrist place in our template is the fact that the “bathtub” model of causation that is often invoked to figuratively represent the scrutinized theoretical understanding of causal mechanism has strong affinities to figure 3, which we develop next and which helps us to scrutinize the (implicit) understandings of causation that sense makers adhere to.

Kurki has reminded us that we can draw on Aristotle for reflecting on the meaning of causation. For a productive use in the current context, we have to transfer Aristotle’s famous **four causes** into the context of the social sciences and translate them into the language of modern social science theory. Aristotle distinguishes four kinds of causes: material, formal, efficient and final causes. According to Kurki, material and formal causes are located on the structural level of analysis, whereas efficient and final causes are located on the lower level of individual or corporate agency. Material causes can be understood as the distribution of material resources and natural conditions that enable and delimit the potential range and direction of action; formal causes can be conceived as the normative-cognitive structures (discourses, frames) which define the possible (imaginable) range and direction of action. Within a social science context, final causes can be understood as purposes which mobilize and motivate action. Finally, efficient causes refer to agents that produce action through their pushing and pulling activities. Figure 3 shows how these four causes can be located within the conceptual space that refers to the ontological aspects that we scrutinized in the two foregoing subsections—a conceptual space that Wendt used for locating theories in the field of International Relations.

The location of these four kinds of causes is only the first step for delineating different understandings of causation that are in line with the goal of providing meaningful explanations of the social world. The second step is to show how a specific way to refer to these generic causes can be coherently aligned to a specific methodological approach.

The strongly holistic approach presumes that material or ideational **structures** (material and formal causes) strongly determine the formation and motivation of individual actors. In

---

34 Brady 2008: 227.
35 e.g. Coleman 1990; Esser 1993; Elster 1998; Hedstrom and Swedberg 1998.
36 Blatter and Haverland 2014, 95-97.
37 As laid out in Blatter and Haverland 2014, 79-143.
38 Kurki 2008.
40 Coleman 1990.
41 Wendt 1999, 29-32.
43 Wendt 1999, 29-32.
44 Coleman 1990.
45 Wendt 1999, 29-32.
46 Wendt 1999, 29-32.
47 Wendt 1999.
A first step in order to move from holism to relationalism is to broaden the range of causes that are taken into account on an equal footing. In this case, we strive for an explanation that not only includes all four causes, but also shows how the four causes mutually constitute, transform or stabilize each other. We reach such a complete picture when we combine a close interaction and communication between the subjects of the study and the social scientists’ theoretical knowledge which leads to a “fusion of horizons.” Therefore, we call the corresponding ideal-typical method “Co-writing Culture.” Ethnographic studies are usually very close to this ideal-type.

A more radical turn away from holism combines internal coherence with external difference and plurality. Such an approach strives for explanations that are internally coherent inasmuch as they combine and specify different causes that correspond to a consistent materialist or idealist understanding of the nature of social reality (scientific paradigm). Insofar, the corresponding ideal-typical method, Congruence Analysis (CON), is fully in line with sense-makers’ striving for coherence. But, in contrast to other sense making understandings of causality, it does not presume that providing orientation demands a single integrated explanation. Instead, it assumes the productivity of a plurality of theories defined as internally coherent and externally distinctive explanatory schemas. A plurality of distinct theories makes it possible to develop multiple and divergent explanations of a phenomenon. Within such a Congruence Analysis, it is important to reflect on the relationship between the different theories. In principle, they can complement or compete with each other. Furthermore, within a CON the divergent theories must be put to a systematic empirical test in order reveal the (divergent) levels of congruence between each causal schema and empirical observations, a feature that brings CON close to truth-seekers’ striving for correspondence (between the explanation and the real world). Sense-makers can use the results of such a Congruence Analysis in order to reflect on the appropriate standing of the theories in the (scientific/social) discourse. Truth-seekers, in contrast, can use the results of a congruence analysis in order to reflect on the truthfulness of the divergent explanations within the specific field of the study to which they have been applied.

Congruence Analysis, like Causal-Process Tracing, is based on a relationalist understanding of causality inasmuch as both methods strive for explanations which include a multiplicity of causal factors on different levels of analysis (structure and agency). In a CON, though, each explanatory schema must be theoretically coherent, whereas CPT is open to the possibility of revealing the working of a combination of social mechanisms in which each single social mechanism is aligned to a distinct theory or paradigm. In other words, CPT is applied in order to create a comprehensive explanatory model that includes a plurality of causal factors, whereas CON develops a plurality of coherent theories in order to produce a comprehensive, in the sense of being multifaceted, understanding of an empirical phenomenon.

---

44 Van Hulst, Blatter, Haverland 2016, xi-xii.
45 See Van Hulst, Blatter, Haverland 2016, x-xi.
### Table 2a: Purposes and Epistemological Principles of the Six Qualitative Methods

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Research Design</th>
<th>Aspired Explanation</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth-Seeking</td>
<td>Which effects do a specific cause (X) have?</td>
<td>=&gt; X-centred research design</td>
<td>Means + preconditions for valid inferences: formal logic, reflections on the similarity of cases</td>
<td>Transparency: selection of variables and cases; scoring</td>
</tr>
<tr>
<td></td>
<td>Which configurations of conditions make a specific kind of outcome (Y) possible?</td>
<td>=&gt; Y-centred research design</td>
<td>Means + preconditions for valid specifications and concatenations: set theory, reflections on optimal calibration, consistency and coverage</td>
<td>Replicability: access to raw data; robustness tests</td>
</tr>
<tr>
<td></td>
<td>Which underlying mechanisms (M) make causes produce an effect?</td>
<td>=&gt; M-centred research design</td>
<td>Means + preconditions for valid specifications and concatenations: empirical density and depth; intensive reflections on relationships among causal factors</td>
<td>Traceability: detailed description of processes; access to original sources</td>
</tr>
<tr>
<td>Sense-Making</td>
<td>Which specific theory (T) provides better/further explanations?</td>
<td>=&gt; T-centred research design</td>
<td>Comparatively high congruence between a coherent theory and a crucial case</td>
<td>Fairness: unbiased selection, specification, and application of theories</td>
</tr>
<tr>
<td></td>
<td>Which interpretative signs and processes characterize a specific culture (C)?</td>
<td>=&gt; C-centred research design</td>
<td>Consistent narrative of cultural practices co-produced by scholar and subject</td>
<td>Authenticity: interpretations are shared/supportied in the field of study</td>
</tr>
<tr>
<td></td>
<td>Which underlying structure (S) stabilizes and/or transforms a social system?</td>
<td>=&gt; S-centred research design</td>
<td>Creative re-construction of social reality based on a comprehensive worldview</td>
<td>Reflexivity: researcher(s) is positioned in the research context</td>
</tr>
</tbody>
</table>

#### Characterizing Different Ideal-Typical Methods

Tables 2a (above) and 2b (next page) summarize the delineated characteristics of the six methods in their ideal-typical form.46

#### Concluding Remarks on the Potential Uses of the Typology

On a practical level, the typology of methods helped us to write the introduction of our latest textbook on qualitative methods in Political Science.47 There, we presented a set of prototypical research goals and questions (as done in the second section of this essay). The core message for practitioners is that each prototypical research question implies a specific configuration of epistemological principles and ontological presumptions with its corresponding methods of data creation and data analysis.

On a methodological level, though, the presumption that is implied in such an advice might be questioned. In research practice and in methodological contributions, we find applications and descriptions of the methods (with the same or similar labels) which do not correspond to our ideal-typical descriptions. For example, many scholars and methodologists pursue Qualitative Comparative Analysis (QCA)—the most common label for the family of methods that correspond to our ideal-type CCA—or Process Tracing (PT) as purely truth-seeking endeavours. Moreover, interpretive scholars rarely put a plurality of theories to empirical scrutiny in the systematic way that is necessary for a good Congruence Analysis. Some scholars will challenge the presumption that the ideal-types presented here represent the best ways to search for valid and reliable answers to the formulated questions. Nevertheless, these ideal-types might challenge, in turn, the presumption that one must be either seeking the truth or making sense. My claim is that if adequately designed, some methods allow for pursuing both at the same time. Furthermore, the framework might be helpful to clarify which distinct strategies we have to apply in the various steps of a research project when we use these methods either as pure truth-seekers, as pure sense-makers, or as scholars who combine truth-seeking and sense-making.48

On the most general level, the framework and typology presented here should stimulate a debate about the meaning of “multi-method research.” Currently, the dominant understanding equates multi-method research with a combination of

---

46 Blatter 2016 contains not only a third table in which the corresponding methods of data collection/creation and data analysis/interpretation are summed up, but also a more comprehensive description of four out of the six ideal-typical methods.

47 Blatter, Langer and Wagemann 2017 12-17.

48 For first attempts to do so in respect to PT and CCA, see Blatter 2016; Blatter and Huber 2017. One important insight has been that Bayesian reasoning has to be applied differently.
### Table 2b: Ontological Presumptions of Six Qualitative Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Configurational Comparative Analysis</th>
<th>Causal-Process Tracing</th>
<th>Congruence Analysis</th>
<th>Co-Writing Culture</th>
<th>Con-Textual Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementarism</td>
<td>All kinds of material and ideational factors understood as variables</td>
<td>All kinds of material or ideational factors understood as conditions</td>
<td>A complete set of causal factors on the macro-level and on the micro-level of analysis</td>
<td>A theoretically coherent set of material or ideational factors on the macro- and micro-level of analysis</td>
<td>A coherent set of material or ideational factors understood as structures</td>
</tr>
</tbody>
</table>
| Holism                | Divergent methods (e.g. statistical analysis and Process Tracing, or within-case analysis and cross-case analysis) aiming at strengthening the internal and/or external validity of results, whereby these results represent the answer to the same research question. The presented typology, instead, highlights the fact that divergent methods complement each other in a much more fundamental way by allowing us to pursue different research goals and to answer quite distinct kinds of research questions.

### References


Blatter, Joachim, Frank Janning, Claudius Wagemann. 2007. Qualita-
Opportunities and Obstacles in Distributed or Crowdsourced Coding

Brian R. Urlacher
University of North Dakota

Scholars interested in translating large quantities of qualitative information into structured datasets have an expanding range of options, each with different strengths and weaknesses to be balanced. Whether a researcher opts for an inductive or deductive strategy for creating conceptual frameworks, in the end, the task of merging concepts with actual data is an “interpretive exercise.” For qualitative data—be it transcripts, field notes, or texts—this exercise has traditionally been carried out by researchers, working alone or in small teams. Yet, relying on human coders can be slow and expensive and can raise questions about inter-coder reliability and validity.

The rise of computerized techniques for coding therefore has tremendous appeal for researchers looking to leverage breakthroughs in natural language processing and qualitative analysis software. As another way to improve speed and overcome some of the limitations of traditional coding methods, researchers are drawn to a middle ground between small teams of human coders and automated systems: The emergence of “micro-task markets,” such as Amazon’s Mechanical Turk platform, promise to reduce the cost and the time associated with human coding while avoiding the technical challenges of natural language processing.

While Mechanical Turk and other strategies for crowdsourcing information collection and processing are beginning to revolutionize the social sciences, there are potential problems associated with distributing research tasks. These problems are just starting to come into focus and are not always well understood. My own experience with “distributed” or crowdsourced coding of data suggests that micro-task markets like Mechanical Turk may pose reliability and validity issues that are more complex and less well understood than those associated with the work of more traditional research teams. This essay begins with an introduction of micro-task markets, specifically the Mechanical Turk system. I then describe an inter-coder reliability and validity assessment I conducted using that platform. I conclude with recommendations for improving reliability and validity when using a distributed coding approach and with a discussion of the strengths and limitations of this approach for coding qualitative data.

Introducing Mechanical Turk

Amazon’s Mechanical Turk platform is a market place that trades in Human Intelligence Tasks (HITs), which the company defines as tasks “that human beings can do much more effectively than computers, such as identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings, or researching data details.” On the surface, the platform is a way to execute complex computational tasks, but behind the scenes human workers provide the intellectual power to make the system operate.6

Since its launch in 2005, Amazon’s Mechanical Turk platform has emerged as an invaluable research tool for scholars in many academic disciplines, to the extent that—over the last ten years—it has become a standard tool of behavioral psychologists and is making inroads into other social science disciplines. Mason and Suri provide a review of the uses of Mechanical Turk by researchers in the behavioral sciences, noting the use of the system for surveys, experiments with random assignment, and experiments or simulations involving interactions between study participants.7 For researchers, the appeal of the Mechanical Turk system is the flexible employment of workers to accomplish a range of tedious tasks, quickly and cheaply.

Most of the scholarly applications of Mechanical Turk use the platform to administer surveys or to carry out experiments. There have, however, been limited efforts to utilize Mechanical Turk in data collection and coding. Preliminary studies suggest that this research application may be viable. For example, Alonso and Mizzaro found the document-coding by experts and workers hired through Mechanical Turk to be of comparable quality.8 In another study, Kittur et al. used Mechanical Turk to code Wikipedia entries.9 The Mechanical Turk worker coding was compared to coding produced by Wikipedia administrators. Kittur et al. found that Mechanical Turk could, in some instances, be used to produce reliable results, but they caution that the structure of the HIT was an

Brian R. Urlacher is Associate Professor of Political Science and Director of International Programs, CoBPA, at the University of North Dakota. He is online at brian.urlacher@business.und.edu and https://und.edu/directory/brian.urlacher.

1 Glaser and Straus 1967.
2 Goertz 2006.
3 Druckman 2005, 258.
4 Schrodt and Van Brackle 2013.
5 Mechanical Turk FAQ 2017.
6 The Mechanical Turk system is a twenty-first century interpretation of a famous eighteenth-century hoax perpetrated by Wolfgang von Kempelen. Von Kempelen’s con involved disguising a chess master within a “mechanical device” named the Mechanical Turk. Von Kempelen convinced the luminaries of Europe that he had invented a machine that had mastered the game of chess (Levitt 2006). Amazon’s Mechanical Turk platform operates on a similar principle (albeit without von Kempelen’s deceit).
7 Mason and Suri 2012.
8 Alonso and Mizzaro 2009.
9 Kittur et al. 2008.
important factor driving the accuracy of the coding.

These initial studies suggest that well designed coding tasks can be effectively crowdsourced. However—in my own experience working with a traditional team of coders on research projects, and as a worker on Mechanical Turk—there is often value in having coders with a base of knowledge on a topic. In a distributed work environment where the identity of workers is unknown, such a foundation of background knowledge cannot be assumed. Furthermore, coders working in more traditional teams have the opportunity to learn and refine their coding over multiple batches of material and through dialogue with other coders. This kind of group and individual norming is almost certain to be absent in the context of distributed coding where coders likely do not discuss their thought processes and may only code a few cases before moving on to other tasks.

Evaluating Distributed Coding

To better understand how to get the most accurate results from distributed coding, I carried out an inter-coder reliability analysis and a validity test on three different groups of Mechanical Turk workers. These groups included workers from the general Mechanical Turk workforce, from a subset of highly proficient and experienced users identified by Amazon as “master” workers, and a third group of non-master workers who completed a brief online training module. The training module consisted of approximately 600 words of background information on negotiations, conflict termination, comprehensive and partial agreements, and third parties. This information was followed by six multiple choice or true false questions. Successful answers to the comprehension quiz were required before coders could access their first case. Coders were paid a $0.50 bonus for completion of the training (twice the rate paid for individual HITs).10

For this analysis, I created a seventy-case test dataset comprised of qualitative descriptions of negotiations to end interstate wars. The descriptions were culled from the Uppsala Conflict Data Program’s (UCDP) conflict encyclopedia.11 The qualitative descriptions ranged from 23 to 551 words with an average length of 140.6 words and provided a general overview of negotiation activities occurring in a given conflictyear. After reviewing the descriptions, I identified three variables relating to the onset and termination of negotiations that I believed could be extracted from the texts. These three variables were structured to provide greater precision than is included in the standard UCDP coding.

The HIT involved coding three variables related to the start of a new negotiation process, the way that negotiations ended, and if a partial agreement was reached in a given year. For the new negotiation variable three responses were possible (New Negotiations, Negotiations Continued from Earlier, and No Negotiations Occurred). For the termination of negotiation variable four responses were possible (Negotiation Continued, Negotiation Collapsed, Negotiation Ended with a Partial Agreement, Negotiation Ended with a Full Agreement). Finally, a Yes or No question was posed to record if partial agreements were reached as part of a continuing processes. For each of these three variables, brief descriptions were provided to coders to explain each category. Selecting some responses prompted coders to also extract dates from the text.

In assessing inter-coder reliability and validity, I used the proportion-of-agreement statistic, which is calculated by converting categorical variables into an array of dichotomous variables (coded as 0 or 1) representing each category. The proportion of agreement statistic is the percent of cases in which two coders (or in this case coding processes) produced identical results (both 0 or both 1) for a given category. Fleiss et al. describe this procedure and note that the proportion of agreement measure is “the simplest and most frequently used index of agreement.”12

The 70 descriptions of interstate negotiations drawn from the UCDP’s conflict encyclopedia were each coded four times (including my own coding) for a total of 280 observations.13 A total of 73 different Mechanical Turk workers participated in coding cases. This included 10 master workers, 17 non-master workers who did not receive training, and 46 non-master workers who completed the training module.14 While this is perhaps fewer cases and coders than might be desired in terms of statistical power, the results in this limited study served the purpose of helping me to assess the viability of re-coding the UCDP’s conflict encyclopedia using Mechanical Turk.

Assessing Reliability

A comparison of the various sets of workers on Mechanical Turk showed disappointing levels of consistency in how the same data was coded. Tables 1A-C provide the percentage-of-agreement scores for the three variables coded by Mechanical Turk workers. While the proportions are generally greater than 0.5, simple randomness would meet that level of agreement for

---

10 It is common for Mechanical Turk workers to be asked to complete training or certification processes before being allowed access to HITs. In my own experience as a Mechanical Turk worker, I found that some certifications were very labor intensive to complete, and that compensation was never offered for work done to complete a certification process.

11 Earlier versions of the UCDP conflict encyclopedia allowed for users to generate custom reports. I selected interstate conflicts with negotiations occurring in a given year. I then requested qualitative descriptions of these negotiations between 1975 and 2012. There were 70 descriptions of negotiations available for analysis.

12 Fleiss et al. 2003.

13 I include my own coding here not as a “correct” measure but rather as a way to evaluate how different pools of researchers performed compared to my own subjective coding of the UCDP descriptions.

14 Master workers completed an average of 7 HITs each. Non-master workers who were not asked to complete training, completed 4.11 tasks on average. The average number completed for workers who were asked to complete the training was 2 HITs, and more than 70% completed only 1 HIT. I suspect that the one-time bonus for completing training may have motivated a number of workers to complete one hit even though they may have been only marginally interested in the actual task.
A dichotomous variable. For a three-category variable, random chance would produce a proportion-of-agreement score of about 0.55. Similarly, a four-category variable would have a baseline expected proportion of agreement of 0.625. For comparison, in projects, in which I have participated and which have used traditional teams of human coders, a threshold of 0.9 (indicating 90% agreement) has often been used as a minimum acceptable level for intercoder reliability.

The first variable (see Table 1A), which related to the onset of negotiations, had proportion-of-agreement scores distributed fairly evenly between the minimum of 50%, which actually falls below the baseline expected by chance, and the more respectable maximum of 91%. The second variable (see Table 1B), which refines how negotiations ended, showed a similar distribution. On the low end, proportion-of-agreement scores approached what would be expected from randomness. On high end, agreement levels reached 95.7%—although most categories fell closer to the average score of 79%.

Table 1C presents the results for the variable asking about negotiations that continued after parties reached a partial settlement. This dichotomous variable produced the lowest proportion of agreement scores with an average of only 62.4%.

Two issues seem to arise from these tables. First, the low inter-coder reliability scores suggest either a high level of error in the coding processes or real and meaningful differences in the performance of the various pools of Mechanical Turk workers. A second issue relates to how my own coding tracked with that of other workers. While master workers generally produced results closest to my own (80.9% on average), coders that completed the training module seemed to be slightly more out of sync with my own coding (75.6 percent on average). This difference is statistically significant and alarming in that the intent of the training was to produce results that would be more comparable to what I would have produced on my own.\footnote{All statistical significance tests presented in this paper are set at the .05 level. Standard errors are calculated using the formula for standard error of a proportion.}

\section*{Recommendations}

It is difficult to know how much of the proportion-of-agreement results are attributable to the distributed coding approach. Intercoder reliability can be affected by the specific individuals that participated in the coding and by the underlying ambiguity in this project, including the coding protocols and data. Without a comparison to a broader set of human coders operating in more traditional teams, it is difficult to know how much the limited learning and coder expertise are responsible for the results.

For sources of error external to a specific project, a reasonable strategy for improving the quality of coded data would be to code each case multiple times and use the modal response. While drawing multiple perspectives may help to even out idiosyncratic error and reduce overall measurement error, there is a practical issue to be considered. Although coding...
each case three or five or ten times may improve results (note that I have not tested this in the context of Mechanical Turk), this solution increases both the costs associated with coding each individual case and requires an additional layer of processing by a researcher to resolve disagreements between coders. It should also be noted that this approach only addresses the problem of random error. The repetition solution rests on the assumption that coders are on average able to generate consistently valid results. It is to this assumption that I now turn.

Assessing Validity

While this project initially sought to produce more fine-grained coding of negotiations than what was available through the UCDP’s coding, two of the categories related to the termination of negotiations run roughly parallel to the UCDP’s own coding. Thus, as a validity check, I compared the various pools of coders against information from the UCDP Peace Agreement Dataset. My own coding of the UCDP descriptions aligns with the Peace Agreement Dataset a little more than 90% of the time. Somewhat unexpectedly, given the findings above, trained coders are statistically comparable in their performance, with proportion-of-agreement scores of 88.5% and 91.4%. This was also statistically better than the scores for both master workers and untrained workers, which hovered around 80%. The differences across groups seems to be driven by the reduction in false positives. That is, trained coders were much less likely to incorrectly record a full or partial agreement based on the descriptions than were master workers or untrained workers.

On the whole, the validity check is somewhat encouraging. In general, Mechanical Turk coders produced results that were not strikingly different from what the UCDP team produced, although the results would not be strong enough to satisfy most researchers. Still, two points are worth noting. First, the brief descriptions of negotiations provided to Mechanical Turk workers are devoid of larger contextual informa-

Table 1B: Proportion of Agreement by Coding Processes

<table>
<thead>
<tr>
<th>General Workers (No Training)</th>
<th>General Workers (With Training)</th>
<th>Master Workers</th>
<th>Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continued</td>
<td>62.8%</td>
<td>75.7%</td>
<td>71.4%</td>
</tr>
<tr>
<td>Collapse</td>
<td>80.0%</td>
<td>81.4%</td>
<td>77.1%</td>
</tr>
<tr>
<td>End with Partial Agreement</td>
<td>84.2%</td>
<td>80.0%</td>
<td>85.7%</td>
</tr>
<tr>
<td>Ended with Successful Negotiations</td>
<td>77.1%</td>
<td>81.4%</td>
<td>81.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Workers (With Training)</th>
<th>General Workers (With Training)</th>
<th>Master Workers</th>
<th>Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continued</td>
<td>------</td>
<td>64.3%</td>
<td>62.3%</td>
</tr>
<tr>
<td>Collapse</td>
<td>------</td>
<td>84.3%</td>
<td>77.1%</td>
</tr>
<tr>
<td>End with Partial Agreement</td>
<td>------</td>
<td>87.1%</td>
<td>95.7%</td>
</tr>
<tr>
<td>Ended with Successful Negotiations</td>
<td>------</td>
<td>75.7%</td>
<td>84.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Master Workers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Continued</td>
<td>------</td>
<td>75.7%</td>
<td></td>
</tr>
<tr>
<td>Collapse</td>
<td>------</td>
<td>84.3%</td>
<td></td>
</tr>
<tr>
<td>End with Partial Agreement</td>
<td>------</td>
<td>88.6%</td>
<td></td>
</tr>
<tr>
<td>Ended with Successful Negotiations</td>
<td>------</td>
<td>82.8%</td>
<td></td>
</tr>
</tbody>
</table>

16 Harbom et al. 2006; Högbladh 2011.
Table 1C: Proportion of Agreement by Coding Processes

<table>
<thead>
<tr>
<th>General Workers (No Training)</th>
<th>General Workers (With Training)</th>
<th>Master Workers</th>
<th>Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Agreement (with negotiations continuing)</td>
<td>55.7%</td>
<td>61.4%</td>
<td>71.4%</td>
</tr>
<tr>
<td>General Workers (With Training)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Agreement (with negotiations continuing)</td>
<td>--------</td>
<td>57.1%</td>
<td>61.4%</td>
</tr>
<tr>
<td>Master Workers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Agreement (with negotiations continuing)</td>
<td>--------</td>
<td></td>
<td>67.1%</td>
</tr>
</tbody>
</table>

Table 2: Evaluation of Coding Processes Relative to UCDP Coding

<table>
<thead>
<tr>
<th>General Workers (No Training)</th>
<th>Proportion of Agreement</th>
<th>False Negatives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCDP Full Agreement</td>
<td>80.0%</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>UCDP Partial Agreement</td>
<td>81.4%</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>General Workers (With Training)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCDP Full Agreement</td>
<td>88.5%</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>UCDP Partial Agreement</td>
<td>91.4%</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Master Workers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Agreement</td>
<td>81.4%</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>UCDP Partial Agreement</td>
<td>81.4%</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Researcher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Agreement</td>
<td>92.80%</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>UCDP Partial Agreement</td>
<td>90.0%</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
tion so it is difficult to know if a more accurate coding is reasonable to expect given the descriptions. Second, there seems to be a strong bias on the part of Mechanical Turk workers toward false positives. That is, they seemed more likely to report that negotiations occurred, when there was little evidence to support this. This is most evident with untrained Mechanical Turk workers, who coded approximately 16.4% of cases as false positives.

**Recommendations**

A striking portion of the error in the validity test resulted from false positives. There is a case to be made that this may be a structural bias built into the Mechanical Turk platform. Many HITs have tests embedded within to ensure that workers are completing tasks dutifully. There are considerable information and power asymmetries between workers and those hiring for tasks. Thus, workers may be conditioned to look for “correct” answers in every HIT and may have an aversion to coding null results.

Fortunately, the training module seems to have effectively countered this tendency. Yet, providing training may have inadvertently limited some of the learning that occurs in coding multiple cases. While master and non-master workers averaged around 3 minutes per HIT, it generally took workers 9 minutes to complete the training and their first task. Consequently, the vast majority of workers who completed the training finished only one HIT. In other words, this relatively brief training may have consumed the time workers were willing to dedicate to the task, thus decreasing opportunities to accumulate experience through practice.

Given my experience with attempting to code qualitative data through Mechanical Turk, I have come to the conclusion that distributed coding may be effective when tasks are straightforward and homogenous, but the Mechanical Turk platform may also create incentives that bias results. This bias can potentially be countered through training but as the reliability analysis shows, researchers should proceed with caution. Thus, McClelland’s cautionary advice in using human coded event data seems to also hold for coding generated through Mechanical Turk: “let the user beware.”

**Concluding Thoughts**

The results of a single study of inter-coder reliability and validity should not serve to legitimize or delegitimize an entire approach to data collection and processing, but the findings presented here do provide an opportunity to pause and reflect. The distributed model of coding is not well understood nor documented, and researchers should proceed with caution. There are multiple practical and ethical considerations when thinking about using traditional approaches to human coding, a crowdsourcing model, or a fully computational approach.

At least part of what researchers will need to consider relates to the nature of the task and the nature of the project. Some tasks are relatively intuitive for humans but may be more difficult to automate. Where these tasks are simple and repetitive, there may be real value in pursuing crowdsourcing. For smaller projects, however, the technical barriers of setting up a crowdsourcing platform or machine learning algorithms may be a prohibitive barrier. Yet, economies of scale make setting up a micro-task market negligible in large projects.

In spite of these limitations, crowdsourced or distributed coding offers one clear advantage over the traditional approach to human coding. A distributed network of coders is more likely to be representative across age, gender, race, and the North-South divide. It therefore allows scholars to avoid the cultural or gender bias that can arise from the lack of diversity in teams of coders. Buhrmester et al. observe that Mechanical Turk workers are not noticeably different than the general internet population and are therefore more representative than what is often found in studies that draw heavily from university communities. Furthermore, the Mechanical Turk pool of workers draws globally, although as Ipeirotis documents, this global workforce is not globally representative as the vast majority of workers are based in either the United States or India.

In short, there are potential advantages and opportunities for researchers seeking to draw on distributed networks of human coders. The speed, cost competitiveness, and representativeness of micro-task marketplaces are genuine advantages. Yet, my own investigation suggests that researchers should proceed carefully, consider the appropriateness of a task, and scrutinize the results of distributed coding before fully committing to crowdsourced data processing.

**References**


---

17 In my own experience as a Mechanical Turk worker, tasks often take much longer than advertised and payment can be withheld without explanation or practical recourse.

18 McClelland 1983.

19 Alker 1998, 224.

20 Buhrmester et al. 2011, 4.

21 Ipeirotis 2010.
Surveying the field in mid-20th century, David Easton observed that political science primarily lacks “a conceptual framework or systematic theory to give meaning, coherence, and direction to ongoing research.”1 In his most prominent work, The Political System, he writes, “In political science there has been little deliberate effort to formulate a conceptual framework for the whole field.”2 The problem he undertakes, then, is that of “defining political science,” and thus to “define the core of the field.”3 This was Easton’s heroic mission.

After that publication, Easton enjoyed phenomenal success within the profession. Elected American Political Science Association president in 1969, studies conducted in the 1970s and 1980s showed that his peers ranked him fourth among the most prominent political scientists from 1945 to 1960, and second most prominent in the period 1960-1970. He ranked seventh among the twenty most cited political scientists in the decade 1970-79.4

Since its publication in 1953, Easton’s “conceptual framework” for political science has conventionally been depicted as a seminal work of “positivistic behavioralism.” He has been cast as a leader in the study of political behavior with an understanding of science as an effort to produce empirical hypotheses that can be falsified or verified. For example, in a 2006 survey of approaches to the study of politics, Mark Bevir classifies Easton as one of the foremost “behavioralists,” and an advocate of behavioralism as an “expression of the turn toward positivism.”5 Bevir then defines the “positivist concept of science” as seeking “universal, deductive, predictive, and verifiable theory.”6 Bevir reflects a well-established categorization of Easton as a “positivist,” and the equation of positivism and behavioralism.

Ironically, despite this wide spread characterization of Easton’s point of view, a question remains as to how well this classic work, The Political System, has been understood. Historian of the political science profession, John Gunnell, has observed that this book “has now become somewhat a prize among the twenty most cited political scientists in the decade 1970-79.”1

Since its publication in 1953, Easton’s “conceptual framework” for political science has conventionally been depicted as a seminal work of “positivistic behavioralism.” He has been cast as a leader in the study of political behavior with an understanding of science as an effort to produce empirical hypotheses that can be falsified or verified. For example, in a 2006 survey of approaches to the study of politics, Mark Bevir classifies Easton as one of the foremost “behavioralists,” and an advocate of behavioralism as an “expression of the turn toward positivism.” Bevir then defines the “positivist concept of science” as seeking “universal, deductive, predictive, and verifiable theory.” Bevir reflects a well-established categorization of Easton as a “positivist,” and the equation of positivism and behavioralism.

Ironically, despite this wide spread characterization of Easton’s point of view, a question remains as to how well this classic work, The Political System, has been understood. Historian of the political science profession, John Gunnell, has observed that this book “has now become somewhat a prisoner of the perspectives that have subsequently informed its interpretive history.” He questions whether the book has ever been truly understood, and notes that, even “after two generations,” the “actual text has never been fully and carefully analyzed.” Therefore, “it is worthwhile reexamining in detail the structure of the argument and determining its place in Easton’s path to the formulation of a systems analysis of political life.”

He adds that given the passage of time since the book’s first publication in 1953, it “is in some respects possible today to understand the book better than it was understood in the 1950s, and it may be possible to understand the author better than he understood himself.”

In this paper, I will undertake the same “reexamining” project as Gunnell, but from an entirely different perspective. This exegeses of the systematic political theory Easton introduces in The Political System focuses not on its history, but on its singularity; that is, its creative origins, intentionality, and interconnectedness. Taking this fresh look at the book has a surprising result. Freeing Easton from, as Gunnell has observed, the prison of “the perspectives that have subsequently informed its interpretive history,” this essay will show, perhaps for the first time, that The Political System actually makes a case for a thoroughly interpretive, as opposed to a positivistic, political science.

The Need for a Systematic Theory of Politics

While Easton advocated developing the study of politics as a science, his understanding of science is different from that embodied in contemporary positivism. He understands the concept broadly as an organized frame of reference with an empirically based “equilibrium theory” for, among other things, being too mechanistic to be a part of his systematic political theory, he is often associated with it. For example, in his history of the Caucus for a New Political Science, Clyde Barrow states that the “central focus” of Easton’s approach is “to understand how decision-making (i.e., authoritative allocations of values) facilitate the equilibrium of the overall social system,” at page 217 (emp. ad.). Barrow 2008, pages 216-217. See Kelleher 2017 for an extensive critique of Barrow’s misrepresentation of Easton.

A more recent instance of conflating Easton with other writers under the rubrics “behavioral political science” and “functionalism” is by Mary Hawkesworth who writes that for Easton, as for these schools of thought, the political system is seen as functioning “so as to maintain homeostatic equilibrium.” See Hawkesworth, “Contending Conceptions of Science and Politics,” page 46, in Yanow 2014. But we will show that, as to Easton, nothing could be further from the truth.

1 Easton 1953, 1971, 52.
4 Lynn 1983; Robey 1982; Roettger 1978.
5 Bevir 2006, quote at page 18 in online version.
6 Gunnell 2013, 198. Besides Bevir, there are many other instances supporting Gunnell’s observation. For instance, although Easton re-
Easton found that one of the consequences for political science of lacking systematic theory was the pervasive condition he called “hyperfactualism.” Rather dramatically he declared that in his day, the “American political scientist is born free but is everywhere in chains, tied to a hyperfactual past.”

Hyperfactualism makes the assumption that “science” consists exclusively of objective and detached fact gathering. But, in Easton’s view, this assumption about the nature of science is way off the mark. He offers a different understanding of science; that is, as an approach by researchers who act not with a blank slate of a mind, but who are infused with and guided by a range of prior interests. In his view, the “aspect of the event selected for description as the facts about this event about the subject. Referring to the literature of his time, Easton regretted what he saw as the “flight from scientific reason, especially in the area of political knowledge.” Indeed, Easton’s theory of the “political knowledge” which political scientists have, and how it is to be verified, may be some of his least appreciated and most unexpected contributions to political science; therefore, they will be clarified in this paper.

Easton envisions science as the apex of human reason. He notes that he is far from the first to search for a science of politics, but he is intent upon pursuing that goal. To have a science of politics, Easton argues, political scientists must begin with a clear idea of what “politics” is. “Where does the political begin and end, and how is it distinguishable if at all, from other kinds of data that we call economic, sociological, psychological, and so on.” Such a theoretical demarcation is needed “to identify the significant variables necessary to explain political activities and to show their interrelations.”

**Below the Level of Theory**

But, without a systematic theory to guide their selections of topics and related facts, how were political science researchers in the past able to write anything about politics? Indeed, how could separate fields of social science, such as political science and psychology, even develop without theoretical guidance?

In Easton’s interpretation, it was a pioneer’s subjective understanding of “the intrinsic logic” of each field in the social sciences that lead to their development as separate areas of research. From his reading of history, Easton had learned that “sciences do not arise capriciously.” “Some process of selection does take place.” Easton notes that specializations in the social sciences have not developed as mere “historical accident,” but by “a rationale of their own.” Out of the complexity of “social life,” leading writers selected clusters of interactions for study. Original thinkers in various fields linked together sets of recurring issues, problems, challenges, observations, and questions, often led by an intuitive sense that the issues were related. Writers who shared overlapping intellectual passions helped develop specialized fields of social science.

For instance, psychology was pulled together, or socially constructed, by writers who shared the intellectual passion to understand individual behavior. By the subject matter they put together they distinguished their field from other social sciences, such as economics, political science, or geography. Easton points out that those writers intuited a kind of “logic of the situation.” Those who made careful studies of particular aspects of society understood and shared certain “key questions” which helped to define their professions. Unarticulated beliefs, and values serve as an interpretive framework that guides research and gives particular facts, out of the whole, their special significance. Without such “theoretical assumptions…it would be impossible for them to select meaningful facts.” Thus, no matter the reasons for the pursuit of research, it is “scientific” if it is conducted with a systematic approach and from within an orderly “frame of reference.”

His concept of science also includes an understanding of verifying or validating the fruits of research, which will be discussed further in a moment. And, as we will see, these principles inform Easton’s theory of political knowledge.
tacit factors, such as a personal sense of the subject and intuition, served as research guides in lieu of a defined conceptual framework. Thus, leading figures paved the way, following their own inner sense of the subject matter.

For Easton, political science in particular, has risen largely out of the personal interest of pioneering researchers in understanding how policy for society is made and implemented. In the absence of a systematic theory defining the field of political science, “most students of political life do feel quite instinctively that research into the political aspects of life does differ from inquiry into any other [aspect of social life].” 26 “Quite instinctively” political science researchers select some kinds of facts rather than others. 27 Among Easton’s most admired pioneering political scientists is V. O. Key, in part because he had a “feel for politics,” which distinguished his writings from the more crude hyperfactualists. 28

Common Sense

Apart from instinct and feel, some political scientists have used “common sense” notions of what is political as their implicit orientation to the subject. Common sense understands politics to include politicking—“maneuvering for position and power.” 29 It also understands politics as “an activity related in some vague way to problems of government or the making of power.” 30 Thus, Easton wants “to raise these assumptions to the point of consciousness for the purposes of careful examination.” 31 Common sense can be misleading or mistaken, and can vary among researchers, so an explicit systematic theory can center collective attention, and critically sift the relevant knowledge, if any, from writings based on differing notions of common sense, intuition, etc. 32

The Axiom

Aspiring to make the field more scientific, Easton asks how the conduct of political science research can be made more rational than by following various forms of preconscious intuition. Many of those who doubt the efficacy of scientific rationality have been relying on the guidance of such nebulous factors as feel, common sense, instinct, etc. But, as noted, this can lead to difficulties in a profession for several reasons, not the least of which is that various writers might not share the same sub-conceptual feelings. Thus, Easton will attempt to liberate the profession from its reliance on vague tacit notions by explici
dly defining the core of the field in a way that both appeals to reason and can be criticized rationally. He begins this effort by presenting his basic axiom for political science.

Familiar to political scientists, it may fairly be stated that “politics,” the subject of political science, is that human behavior undertaken in relation to the authoritative allocation of values for a society. 33 This axiom is meant to, among other things, “sum up our common sense conception of politics.” 34 Once its key terms are defined, which Easton does, the axiom raises the vagaries of undefined intellectual sensitivities to a clear conceptual level. Because it is in an articulated form, practitioners in the field will be able to rationally criticize the axiom. As we will see, Easton understood his axiom as having implications for not only the scope, but also the methods of the field. 35

Thus, Easton’s definition of “politics” is a theoretical position, or interpretive framework, and, as such, a way of seeing social interactions, and a principle for organizing and making sense of social observations. From the axiom-guided observations of the complex behavior of individual humans on the ground, which constitute political life, the vision of a persistent political system emerges. As Easton writes, with the axiom guiding observation it becomes clear that “political life constitutes a concrete political system which is an aspect of the whole social system.” 36 This insight calls attention to another fault of hyperfactualism, namely that it conceals “from students of political life the need to view the political system as a whole.” 37

For Easton, such terms as “political life” and “political process” are general references to the same subject matter as the “political system;” namely, all those human interactions done in relation to the authoritative allocation of values for a society. The “political system” entails “all those kinds of activities involved in the formulation and execution of social policy;” i.e., “the policy-making process.” 38 In practice, these specific components are implicated by the axiom. That is, the axiom guides the attention of the political scientist to those components. Because observers know what to look at, i.e., “politics,” they will soon see far more than what the axiom states, as a political system unfolds before them. 39

31 Easton 1953, 1971, 78.
32 Easton 1953, 1971, 129.
33 Easton 1953, 1971, 97.
36 Easton advised that “at a high level of abstraction,” such as his axiom, “theory needs to be free to develop unhampered by excessive worries of verification.” Easton 1953, 1971, 315. This statement alone greatly distinguishes Easton from positivism, which, as we will discuss further below, insists that theories be verifiable.
40 Some readers of The Political System might find it ironic that given its title, in the book, Easton does not elaborate specifically on the five components of his theory (inputs, conversion, outputs, feedback, and environment). Instead, he carries out an extended presentation of that theory in two succeeding works, A Framework for Political Analysis, Easton 1965a. And A Systems Analysis of Political Life.
Political Life as Human Life

Throughout his discussion, Easton stated his axiom in slightly different variations, and also mentioned some qualifying nuances he intended but which are not always specified in his definition of politics. For example, he notes that out of all the forms of social activity some are “closely related to what we call political life.”41 Indeed, one key element of Easton’s axiom is that it contemplates “life;” specifically, “the political aspects of life,” or “the political side of life.”42 By “life,” of course, he means human life. He observes, for instance, that “What we have in the concrete social world is a series of events in which human beings are involved.”43 He also wrote, “As a social system, a society is a special kind of human grouping the members of which continually interact with one another and in the process develop a sense of belonging together [or,] common consciousness.”44 He notes that “we are human beings who live in an organized society.”45 While this may seem too obvious to merit mentioning, as we will see, his stress on human life has important methodological and epistemological implications.46

The Axiom as a Rule of Relevance

Guided by his axiom, Easton writes that “Political life concerns all those varieties of activity that influence significantly the kind of authoritative policy adopted for a society and the way it is put into practice. We are said to be participating in political life when our activity relates in some way to the making and execution of policy for society.”47 For him, then, a phrase like “relates in some way to the making and execution of policy for society” becomes a normative criteria, or rule, by which political scientists can make crucial research decisions. He writes, “If the object of [systematic] theory is to identify all the important variables, some criteria are required to determine relevance or importance…Without some guide to the investigator to indicate when a variable is politically relevant, social life would simply be an incoherent wilderness of activities.”48 He hoped that his axiom, and the framework implied by it, would help researchers to cultivate a “keen sense of where and how to look for the locus of power and its influence.”49

Intimacy and Method

Easton is offering political science a more humanized form of behavioralism than the positivistic brand. His behavioralism envisions a distinct conception of the relationship between the researcher and the subject matter. He rejects the impersonal, if not alienated, relationship implicated in positivistic efforts at “objective” description. As we will see, Easton’s method is more empathic. This is why the question of how to obtain “intimate knowledge” of political actors arises in his approach, whereas such an inquiry is immaterial, if not inconceivable, in the positivistic approach dominant in his time (and ours).

Easton argued that his method would be able to “provide the basis for the kind of understanding of their data that students of political life seek.”50 He felt that this framework would further provide such matters as “a clear perspective on the fundamental problems of the logic behind scientific method, unambiguous terminology, the introduction of new techniques and a deep awareness of the need to seek out intimacy with observed phenomena.”51

Easton notes that political science is currently behind the other social sciences in its use of the “repertoire of techniques for controlled observation, such as the varieties of highly developed forms of interview and objective participation, the correlation of data, experimentation, and the testing of theories.”52 Clearly, he anticipated that researchers would draw upon the wide variety of methods available in all the social sciences. For, it is not the method but the axiom that defines the field for Easton. He laments that the use of mere fact gathering “techniques has had the secondary result of keeping the research student from intimate contact with his [or her] material.”53

Easton rejected those more positivistic orientations that envision a mechanistic theory of human action as a series of learned, or programmed, reactions to stimuli. This robotic view eschews empathy and blinds itself to the uniqueness of each individual, and to the role of meaning and volition in behavior. He found unacceptable “the damaging effects of this lack of intimate knowledge about political activity on the products of research.”54 That “damage” is, of course, the lack of human understanding needed to solve some of the most pressing political problems of the day.55

Also, in Easton’s methodology, the researcher can approach the subject matter from different perspectives, such as

---

41 Easton 1953, 1971, 97.
42 Easton 1953, 1971, 96, 126.
46 As to why chimpanzee “politics” do not rise to the level of a political system, see Kelleher, “Can Chimpanzee Politics Constitute a Political System?”
47 Easton 1953, 1971, 128.
49 Easton 1953, 1971, 52.
51 Easton 1953, 1971, 52, emp. ad.
52 Easton 1953, 1971, 49.
53 Easton 1953, 1971, 49.
54 Easton 1953, 1971, 49.
55 Easton 1953, 1971, 49.
Thus, even large-N studies require an element of empathy. People make no sense unless their shared meanings are known. By the political intentions of their members. Mere numbers of political science. Political groups, therefore, are defined first physically movements, determine which crowd is a fit subject for a rally. The meanings and intentions of the actors, less than their behavior's relevance. Whether an actor defines himself or herself as acting in a politically relevant way is a matter for the observer's act of empathic interpretation. What Easton adds includes far more than what the statisticians and mathematicians would include. While important, their brand of abstract knowledge is unable to achieve the kind of intimacy Easton envisions.

The Axiom Compels Interpretive Methods

Easton's axiom not just implies, but compels the inclusion of qualitative/interpretive methods in political science. If political science is the study of behavior undertaken in relation to politics, then the intentions of the actor are determinative of the behavior's relevance. Whether an actor defines himself or herself as acting in a politically relevant way is a matter for the political scientist to determine through empathically observing, or interview, or some related technique. The behavior of a crowd at a football game can be similar to that of a crowd at a political rally. The meanings and intentions of the actors, less than their physical movements, determine which crowd is a fit subject for political science. Political groups, therefore, are defined first by the political intentions of their members. Mere numbers of people make no sense unless their shared meanings are known. Thus, even large-N studies require an element of empathy.

How is a potential voting block to be known, except by the understanding of the shared intentions and meanings that relate the individual persons to one another? What determines the relevance to political science as between a group of persons sitting in a living room for a Tupperware party, or a baby shower, and one gathered for a candidate fundraiser? It is their respective intentions and shared meanings. As Easton notes, observing behavior is not enough, the political scientist "must also be prepared to show what makes it political." It's the intentions.

Two Kinds of Interpretation—Theoretic and Empathic

Although he did not make the distinction explicitly, Easton's axiom-centered systematic political theory appears to imply two different kinds of interpretation. One can be understood as theoretic interpretation, and the other as empathic interpretation. The theoretic interpretation entails the act of integrating data (or smaller sets of meanings) into a pre-existing more general conceptual framework. An example of this is classifying observed behavior according to the categories of Easton's theory of the political system—inputs, outputs, etc. The theory tells the political science researcher what to look for on the ground. A description of a particular pattern of behavior, such as the actions of legislators, is a set of meanings, which can then be subsumed under the appropriate category; the conversion process, for instance. This interpretation gives the behavior technical, or theoretic, meaning for the political scientist.

Positivistic political science also interprets events. It has a theoretical framework into which its so-called descriptions of behavior are integrated. The result, as we have argued, is a portrayal of a mechanistic political behavior. In this view, behavior is moved, but is not self-moving by people acting on meanings. No empathy is needed.

The empathic interpretation suggested by Easton, adds a second step of interpretation. That is, a researcher articulates the meanings that persons and groups of persons are acting upon in relation to politics. This set of meanings can include their definitions of the situations, their ideologies, their goals, strategies, etc. Such an interpretation is constructed by the researcher in conjunction with the subjects. And that entails a personal identification of the researcher with the subjects more broadly as fellow humans. This is the way intimate knowledge is obtained.  

The logical operation of theoretic interpretation is more technical than that of empathic interpretation, which is more intuitive. But both types of interpretation, once given in written form, can be rationally criticized, and thus rejected, corrected, or confirmed by other members of the profession. To

---

60 Easton 1953, 1971, 203.
61 Easton 1953, 1971, 204-205.
63 Easton 1953, 1971, 205.
64 Easton 1953, 1971, 201.
65 Easton 1953, 1971, 49.
some extent, the two intellectual operations must be done together. That is, the intentions of the actors must be empathically interpreted as well as their overt behavior described as, for example, of the legislative sort, in order to be classified within the appropriate category, e.g., the conversion process.

Conclusion

Easton’s axiom based systematic political theory, which includes his theory of the political system, stands as a complete alternative to the positivistic framework. That axiom may now fairly stated to define “politics” as the volitilional behavior of persons and groups of people who are intentionally engaged in matters related to the authoritative allocation of values for society. 64

The methodological principle, that the behavior under study is that of “persons,” necessitates the rejection of positivism’s mechanistic and deterministic form of explanation. Easton’s assumption that humans create the meanings by which they define the political situations they are in, and then act upon those meanings, explodes the old positivistic notion of personally detached “objective” observations of behavior, and compels the position of personally involved empathic observation and interpretation.

Under Easton’s principles, then, the explanation of behavior necessarily entails an exposition of the observer’s understanding of the reasons the actors had for their behavior. In some cases, groups and individuals may not fully understand their own motivations, or even misrepresent them. The political scientist can empathically observe and then clarify these motivations. The idea of causation remains, but the mechanistic tendency is factored out by the axiom’s implication of people as volitional, creative, intelligent, meaning making sentient beings. This soft theory of causation recognizes that no matter what was done, it could have been otherwise. In addition, the hope for a way to verify or refute hypotheses, in the manner of laboratory physics, must give way to a requirement more like the intersubjective confirmation or criticism of findings by peers. No one can, for instance, replicate in a laboratory a political system, working within a shared interpretive framework, and drawing from a plurality of methods. 65

64 Understanding the behavior as volitional adds to the challenge of explaining how a political system is able to persist. Every day, large numbers of individuals and groups act in ways that preserve their political system, although at any time they could choose to do otherwise, as they do in revolutions.

65 See Polanyi 1946, 1958. An activist in the paradigm politics of his day, and apparently in an effort to maintain his alliance with his contemporary “behavioralists,” Easton introduced what he called the “behavioralist credo” in Easton 1965a, page 7. He identified himself with the term “behavioralist,” but he understood he had vast differences with his allies. As he commented following the Credo, “No single way of characterizing [these postulates] is satisfactory to ev-

References


______, 2017b. “How They are Healing Physics Envy in Biology.” Draft online at,https://www.academia.edu/19807254/How_They_are_Healing_Physics_Envy_in_Biology
Introduction to the Symposium

Tim Büthe
Hochschule für Politik at the Technical University of Munich, Germany & Duke University

Trade, Expectations, War and Peace

Dale Copeland’s *Economic Interdependence and War*—discussed in this symposium primarily for its methodology—addresses a classic yet very timely question: Do commercial ties between countries increase or decrease the likelihood of militarized conflict? To answer this question, the book develops nuanced arguments about the conditions under which economic interdependence is more likely to become a force for peace and under what conditions it is more likely to exacerbate the risk of a militarized escalation of a conflict of interest.¹

As the magisterial conclusion of a twenty-year research agenda,² the book also presents a wealth of often meticulous empirical analyses of virtually “all the main cases” of great power conflict, starting in 1790.³ Forty such conflicts are analyzed in qualitative case studies in chapters 3-8 of the book—complemented in chapter 2 by a review of several large-N quantitative studies of the relationship between interdependence and war, which Copeland revisits to highlight what their findings suggest (and what they might have missed) about the importance of economic considerations in great power conflicts.

*Economic Interdependence and War* is a highly ambitious book, which sets out to “resolve,” once and for all, multiple debates over the relationship between economic interdependence and war (some of which, as Erik Gartzke points out in his essay for this symposium, go back hundreds of years) by “answer[ing] most of the outstanding questions surrounding the issue.”⁴ That’s a tall order.

The theoretical core of the book’s answer to those outstanding questions is what Copeland calls “Trade Expectations Theory.”⁵ It argues, in a nutshell, that as long as a country’s leaders expect continued access to foreign markets for important inputs and for the country’s exports, economic interdependence will reduce the likelihood of militarized conflict by making such an escalation more costly. This part of the argument assumes that trade is economically beneficial for both sides, that militarized conflict interrupts or at least reduces and threatens such economic exchange, and that the desire to continue to reap the benefits of flourishing trade creates an incentive to maintain peaceful relations. Crucially, Copeland goes beyond classic liberalism by stipulating that this incentive to settle conflicts of interests peacefully operates only as long as the political leadership expects that the country’s trade/economic partners also want to maintain that relationship. This condition is critical because it ensures that, if a country establishes or maintains economic openness in the present, the net present value of the future consequences is positive.⁶

By contrast, if this condition does not hold, greater economic interdependence may actually increase the risk of militarized escalation, Copeland argues. Why might expectations take such a pessimistic turn? And why would expecting a future loss of market access push countries toward war? Starting out under autarky, the correct anticipation of a future loss of market access⁷ should result in forgoing economic interdependence in the first place. Economic openness therefore can coincide with the expectation of declining access to foreign markets for imports and exports only in the aftermath of an exogenous shock to a previously held expectation of ongoing market access. And yet, Copeland cautions, such a turn of events is not uncommon. It may arise, for instance, from a change in any great power’s government “due to elections, coups, domestic instability, and the like.”⁸ whenever such a change is unanticipated and puts in power political leaders with a differing set of preferences.

A change that removes or reduces the prospect of future gains from trade also removes or reduces the above-mentioned

---

¹ Copeland 2015, 1.
² See Copeland 1996.
³ Copeland 2015, 2f.
⁴ Copeland 2015, 1.
⁵ Spelled out in detail in chapter 1 of the book (esp. pp.27-47).
⁶ Copeland 2015, esp. 33ff.
⁷ And expectations are in Copeland’s theoretical model assumed to be always correct in the sense of making optimal use of all available information in light of instrumentally rational pursuit of the political leaders’ national security maximization preferences.
⁸ Copeland 2015, 41; see also 43-46.
constraint. Consequently, if a conflict of interest were to arise for unrelated reasons, and one side were tempted to use military force to gain an advantage in that conflict, contemporary economic openness would no longer provide a counterweight against such temptations. In fact, by creating mutual but asymmetric dependency, economic openness might create just the power imbalance that heightens the temptation to threaten or use military force.

And it gets worse: Economic openness is economically beneficial because it allows for specialization, not just due to static comparative advantage of the kind David Ricardo and his contemporaries already recognized, but also dynamically, for instance, to achieve economies of scale. Once dynamic specialization takes hold, however, it is costly (and in the short run might be impossible) to undo. To sever a relationship of economic interdependence therefore does not mean simply going back to the status quo; it means losses beyond the lost gains from trade. The sudden anticipation of a future in which, due to the expected loss of market access, a country expects to be worse off than it would have been had it maintained autarky, can push the country that (due to unequal interdependence) expects substantially greater losses toward using military force to prevent the other great power from closing or blocking those markets. Copeland thus anticipates military escalation, not just as a way to overcome the intensified feeling of vulnerability, but also to prevent the realization of a (suddenly threatened) negative net present value of economic openness.

What’s At Stake? Theory & Policy

There is much at stake here for IR theory: Michael Hiscox “resolved” the long-standing debate between proponents of the Heckscher-Ohlin/Stolper-Samuelson model and proponents of the Ricardo-Viner/specific factors model of international trade by pointing out that the differing assumptions about mobile vs. specific factors may be more fruitfully treated as scope conditions for the two models. When they are, the seeming incommensurability of the two models drops out, and Hiscox finds strong empirical support for both schools of thought—for each under the conditions when it should indeed apply. As Sherry Zaks points out in her contribution to this symposium, Copeland’s analytical framework might be similarly read as an attempt to resolve long-standing disputes between “liberal” and IR-“Realist” scholars over the effect of interdependence on interstate conflict. In such a reading, the book might then be said to offer an additional, fruitful way of overcoming what Joseph Grieco calls the “schools-of-thought problem in IR,” which would also overcome some of the concerns raised by Timothy McKeown about the conceptualization of expectations. Note, however, that Copeland rejects this reading of the book: As he explains in his rejoinder essay, he believes scholars should treat his Trade Expectations (TE) Theory as a sui generis alternative to both liberal and Realist IR theory.

In addition, there is also much at stake for public and foreign policy at a time when the Trump administration is increasingly behaving like an overstretched declining hegemon in Gilpin’s classic War and Change. Copeland concurs with Gilpin that such declining great powers are generally prone to escalating conflicts militarily—for instance in an attempt to halt their relative decline through a “preventive” war. At the same time, as long as the declining power remains the less dependent country in the interdependent economic relationship (as President Trump tells us is the case for the United States in virtually all of its foreign economic relationships), Copeland expects that the leader’s “awareness of the deleterious effects of the trade-security spiral” give such a leader sufficient “reason to avoid overly provocative policies.” We should of course keep in mind that the current U.S. administration’s portrayal of the United States as being in long-term economic and military decline, might be a tactical move for domestic political purposes, rather than something President Trump and his policy advisors actually believe. But after reading Copeland’s book, the stakes appear even higher when trying to figure out what the most influential members of the current U.S. administration truly believe.

What’s At Stake Methodologically?

Important as the above considerations may be, we selected this book for a QMMR symposium not primarily for its contribution to IR theory nor its policy implications. Rather, we sought to encourage discussion of the big methodological claims and contributions of the book. Most of the major methodological issues that Copeland addresses in Economic Interdependence and War also arise when studying various other subjects across all the empirical subfields of political science. The book, moreover, makes a number of broad claims, such as the claim that qualitative research involves a fundamentally different kind of causal inference because qualitative research (or at least qualitative analyses of individual-level decisions) allow for observing causality “per se.” Such sweeping claims

9 Copeland 2015, 2, 10f, 35f. Note that the underlying conflict of interest, which might escalate or be peacefully resolved, may arise over completely unrelated matters or be itself prompted by economic interdependence. By not taking a position on this issue, Copeland implicitly suggests that TE Theory is meant to capture both possibilities.

10 Hiscox 2002.


12 Applying Copeland’s argument to particular cases and trying to derive policy implications from the book’s findings underscores the importance of addressing the additional conceptual question of what notion of power underlies the claims about increasing/rising and declining powers—which, however, is beyond this symposium.

13 Copeland 2015, 48.

14 Another, even more disconcerting reading of contemporary U.S. foreign economic policy—which one might derive from Copeland’s analysis even if he himself does not specifically suggests it—is that President Trump is trying to sever U.S. economic ties to other countries because he either intends to go to war against some of them or at least wants to restore the ability to credibly contemplate and threaten war to the President’s political tool kit, having already imposed the trade-related costs on the U.S. economy, which might otherwise constrain a President considering military escalation.

15 Copeland 2015, 51; emphasis added. Copeland reiterates this
which the book delivers with great conviction—warrant careful reflection and scrutiny by QMRR scholars, as offered, e.g., by Timothy McKeown in this symposium, before they are accepted or rejected.\textsuperscript{16}

One of the book’s important contributions is its persistent emphasis on causal mechanisms,\textsuperscript{17} followed by detailed empirical analyses of an impressive array of cases of great power conflict, spanning 200+ years and ranging from the military escalation of the conflict(s) between post-revolutionary France and the major European powers (Britain, Austria, Prussia), 1790-1801, to the de-escalation (and ultimately the end) of the Cold War between the United States and Soviet Union in the late 1980\textquotesingle s through 1991.\textsuperscript{18} Copeland stipulates that political scientists should focus more on causal mechanisms—in both their theoretical and their empirical work—and he strongly argues for the superiority of qualitative, causal mechanism-focused research based on archival research (whenever practically possible) over large-N, quantitative, statistical analyses.\textsuperscript{19} Consistent with these stipulations, Copeland presents, for each of his forty cases, information not just about trade levels, political leaders’ expectations regarding their countries’ trading relationships, and other initial conditions of the relationship between the major powers in question. He also presents what Collier and Brady have called “causal process observations”\textsuperscript{20}—in particular, information about the extent to which trade and economic openness were considered by political leaders in their internal deliberations during major power conflicts (including stable or changing expectations regarding the future development of trading relationships).

Why this focus on causal process or mechanisms? Copeland considers it “self-evident”\textsuperscript{21} that rare events in international relations, such as great power conflicts that entail a real risk of military escalation, are “rare for a reason. Rare events in international relations are typically situations where a complex set of factors must come together for the event to occur,”\textsuperscript{22} i.e., they are characterized by the kind of complex conjunctural causality underscored by J. L. Mackie’s classic notion of causality.\textsuperscript{23} And due to this complexity, he submits, there cannot be a single set or “bundle” of conditions that can explain all occurrences of such a rare event. Instead, there must be multiple possible sets of conditions, i.e., multiple “causal pathways,” that result in the actual occurrence of the rare event.\textsuperscript{24} It is not clear why either of these claims should generally hold for rare events, but Copeland is surely right that for phenomena that are characterized by complex conjunctural causality and multiple causal pathways, an in-depth analysis of the causal mechanism is needed to establish a causal account. Extrapolating from the average effect observed for a subset of observations is not possible when these conditions hold.\textsuperscript{25} And studying a phenomenon that is rare might indeed make it practically feasible to study all empirically observed instances of the phenomenon within temporal scope conditions.\textsuperscript{26}

While calling for more careful attention to causal mechanisms in IR scholarship, Copeland also notes that focusing on mechanisms entails potentially severe epistemological risks. Specifically, Copeland criticizes previous qualitative work on interdependence and war (and causal process-tracing case study research in general) for its pervasive selection bias and, consequently, lack of external validity or “generalizability.”\textsuperscript{27} Focusing on causal mechanism, he cautions, tends to bias “qualitative researchers” toward presenting “cases that are particularly useful in illustrating the way these causal mechanisms work in practice” while omitting “cases that do not fit the model.” As a consequence, we may, upon reading such studies, “have confidence” that the hypothesized causal mechanism indeed was at work “for the cases selected, but we have no way of knowing”\textsuperscript{28} whether this finding also holds “for the broader population of cases.”\textsuperscript{29}

In Economic Interdependence and War, Copeland strives to overcome both problems (selection bias and generalizability) by conducting in-depth studies of what he considers “essentially the universe” of cases\textsuperscript{30}—made possible by studying “rare events.” This raises the question: What exactly is, as a methodological prescription, supposed to be rare? The initial conditions, such as a conflict of interests among great powers that trade with each other? Or the extreme outcome of great power war? Or “just” a significant change in the level of conflict (e.g., a shift from escalation to de-escalation)? And what

\textsuperscript{16} On this point, see also Brady’s (2013) and Elman’s (2013) essays on Goertz and Mahoney’s ‘Tale of Two Cultures’ (2012), as well as King, Keohane and Verba’s earlier discussion of the “fundamental problem of causality” (1994, 75f; see also Pearl 2009).

\textsuperscript{17} Copeland 2015, 1f, 13f, 16-18, 43-50, 69f, 72f.

\textsuperscript{18} For each case, Copeland provides something like an historical narrative in the sense of Büthe (2002). His rationale for the book’s overarching historical scope (1790-1991) is discussed on pp.76f; an overview of the cases is provided on pp.79-93.

\textsuperscript{19} For each case, Copeland provides something like an historical narrative in the sense of Büthe (2002). His rationale for the book’s overarching historical scope (1790-1991) is discussed on pp.76f; an overview of the cases is provided on pp.79-93.

\textsuperscript{20} See Linz and Stepan’s (1996) impressively comprehensive study of the fate of third-wave democracies during the early post-transition years for another example of such a study.

\textsuperscript{21} Copeland 2015, 71; see also p.53.

\textsuperscript{22} Copeland 2015, 77.

\textsuperscript{23} See also Büthe 2014.

\textsuperscript{24} See Linz and Stepan’s (1996) impressively comprehensive study of the fate of third-wave democracies during the early post-transition years for another example of such a study.

\textsuperscript{25} Copeland 2015, 13, 53, 70f, 75.

\textsuperscript{26} Copeland 2015, 70.

\textsuperscript{27} Copeland 2015, 70.
exactly is, as a matter of the empirical, historical record, rare about the cases or episodes examined in this book? These questions must be answered clearly to allow the reader to understand: Forty cases of what? And as both Zaks and Gartzke discuss in their respective essays, the answers have profound implications for the extent to which this book, as comprehensive as it is, indeed entails an analysis of something approaching the universe of cases. The answers also tell us to what extent the book’s research design indeed is free from selection bias (granting for the sake of argument the controversial claim that studying the universe of empirically observed cases solves the problems of selection bias and generalizability). Indeed, Zaks and Gartzke are skeptical of Copeland’s assessment that this book has overcome the problem of selection bias, precisely because they understand Copeland’s “true” *explanandum* to be how major powers relate to each other. And this outcome can range from highly conflictual (and rapidly changing) to stably peaceful. Zaks hence suggests that Copeland has selected out “non-conflict interactions” and therefore likely suffers from omission-induced selection bias.

### The Symposium

In order to make this symposium maximally accessible to the broad readership of *QMMR*, including those who might not yet have read the 489-page book, we asked Dale Copeland to provide, as the opening essay for the symposium, a brief summary of what he today considers the key methodological claims and contributions of the book. Copeland’s summary statement is followed by three reviews offering constructive critiques:

**Timothy McKeown** commends Copeland for advancing the theoretical debate, then critically highlights three methodological aspects of the book. He first discusses Copeland’s critiques of large-N statistical studies, especially the claim that quantitative methods do not support causal inference. Second, he examines Copeland’s arguments about the distinctiveness of case studies. Finally, McKeown explores Copeland’s notion of expectations (as the core concept of TE Theory) and examines how we might establish the changing expectations of political leaders.

**Sherry Zaks** approaches the book from a macro-methodological perspective. Informed by the framework developed in her recent article in *Political Analysis*, she asks how exactly Copeland’s theoretical approach is meant to be related to the neorealist and liberal schools of thought in IR. She finds that the book provides an ambiguous answer. This is critically important methodologically because, she argues, the suitability of different empirical strategies depends upon whether theoretical claims are coincident, congruent, inclusive, or truly competing. Zaks also examines Copeland’s approach to rare events and case selection, as noted above.

**Erik Gartzke** joins the conversation as an IR scholar who has himself contributed to the substantive literature on economic interdependence and war. He also comments as a methodologist who has written about analyzing rare events using advanced statistical methods. Viewing TE-Theory’s behavioral assumptions of political leaders, grounded as they are in a probabilistic cost-benefit calculation, as a straightforward extension of classic liberal IR theory, Gartzke commends Copeland for the ambition of reconciling or “synthesizing” liberal and Realist approaches to International Relations. In the end, however, he is not convinced, due to lingering questions about endogeneity (which Gartzke, as many in econometricians, seems to understand more broadly than Copeland) and what he considers incommensurate understandings of “conflict causation” (i.e., why conflicts arise and escalate) in the liberal and Realist tradition. Gartzke also takes issue with the book’s methodological critiques and the alternative approach Copeland advocates, questioning whether we can draw general causal inferences from case studies when the explanatory theory informing them employs a probabilistic notion of causation.

In his rejoinder, Dale Copeland addresses a number of the specific critiques, but above all, he sketches an agenda for future work, viewed through the lens of a new “Approach to Mixed Methods” (AMM). Note that in this AMM framework, the hypothesized causal explanation (e.g., the importance of individual decisions for the international-level outcome) is baked into the researcher’s methodological choices. This implies that it may be impossible for one scholar to fully evaluate more than one (type of) argument—which is useful reminder that science and scientific progress is ultimately a collective enterprise.

### References


---

30 Arguably, it solves neither problem entirely, because the set of observable cases is generally already biased (as Frankfurt School critical theory scholars have long pointed out) and because the inferential logic underpinning causal attribution in process tracing work usually does not provide a justification for out-of-sample extrapolation, unless the possibility of multiple causal pathways can be reliably ruled out (though cf. Büthe 2002 for an argument about probabilistically increased confidence in generalizability under the assumption that multiple causal pathways can be ruled out).


32 His own, substantive closely related research prompts Gartzke to caution that the total cost of war, even for the “winning” side, is so high that, if political leaders are in fact engaging in a rational cost-benefit calculation with good foresight, the potential gains from continued trade (or losses from interrupted trade) should rarely be sufficient to tilt the balance.

33 Even when the motivation for the cost-benefit calculation is maximizing national security, as it is for Copeland (see 2015, 6, 29), since Copeland also assumes high fungibility of economic resources for military ends.
Economic Interdependence and War argues that dependent great powers may be inclined either toward peace or toward actions that can lead to war depending on whether their expectations of the future commercial environment are positive or negative.¹ States that are optimistic about their ability to access raw materials, investment, and markets will be inclined toward moderate policies that build their long-term economic power and that avoid pushing other states into restrictive policies that set off destabilizing trade-security spirals (e.g., China 1985-2015). By contrast, great powers that believe others are cutting them off from access to trade and investment will fear a decline in their power and thus be more likely to initiate military policies that prevent this decline through increased control over economic spheres of influence (e.g., Japan 1930-41). Bridging the divide between liberalism and realism, the book thus seeks to show under what conditions economic interdependence can lead to changes in expectations of future commerce, and thus either to stable international systems or to ones that experience cold and hot wars.

The book offers two main ways to test its propositions: quantitative and qualitative. The first half of chapter two reviews the “greatest hits” of the large-N quantitative work on interdependence and war to show that an expectational approach can explain a number of the anomalies in the current literature. The main focus of my book, however, is on qualitative historical tests of trade expectations theory vis-à-vis commercial liberalism and economic realism. Large-N quantitative methodologies are inherently limited when it comes to exploring the role of leader expectations in great power politics. Because there are no surveys of leaders’ attitudes and perceptions across historical time, quantitative researchers are forced to use indirect measures of expectations—for example, the trend lines of the past three or four years of trade data that leaders are presumed to extrapolate out into the future. Such measures are clearly second-best when there are available documents that reveal how leaders were actually thinking about the future commerce, and thus either to stable international systems or to ones that experience cold and hot wars. Moreover, when dealing with rare events, quantitative research is also constrained in its ability to establish exactly what causal role particular variables might have played in the mix of factors that led to individual cases of war or the ending of rivalry

¹ Copeland 2015.

Dale C. Copeland
University of Virginia

across time.2

The book thus starts its empirical chapters with a short overview of the quantitative research to provide a useful “first cut” test of the possible explanatory value of trade expectations theory.3 The vast bulk of the study is then devoted to an extended analysis of the documents and the best secondary sources for 40 “case periods” from 1789 to 1991. These are periods that essentially cover the universe of wars and major crises between great powers during that timeframe as well as dramatic shifts away from conflict such as the Russian-American détentes of the early 1970s and the late 1980s. Because this is ultimately a study of rare events—great power wars or behaviors that dramatically change the probability of such wars—there is a necessary historical focus on the events themselves and the reasoning behind leaders’ decisions to provoke them.4 Yet, to minimize selecting on the dependent variable, the years prior to key shifts in behavior are also examined. This allows us to see to what extent the planning for conflict, the levels of tension, and the probabilities of war changed as the core independent variables of the competing theories changed.5

The methodological approach of the book is underpinned by three main claims: first, that rare events in great power politics are rare for a reason; second, that we must go beyond merely showing that a factor was present to show what causal role it was playing in the “mix” of forces that led to the rare event; and finally, that the real purpose of good qualitative testing is to establish not simply whether a variable “matters,” but rather how often it matters, in what way, and under what conditions. None of these claims, I believe, should be terribly controversial. But when put together in a coherent way, they offer a potentially distinctive approach to thinking about qualitative testing in international relations. For the rest of this short overview, I will briefly summarize each of these claims.

Great power wars and the crises that raise the risk of war are rare events in international politics for one main reason: they are events of complex conjunctural causality, where a set of factors A, B, and C must come together simultaneously for the event E to occur. Each of the factors is a necessary condition in the mix, and yet when they combine, they become sufficient to produce event E. This “individually necessary, jointly sufficient” (INJS) logic is itself not enough, since for almost all phenomena in international relations, there will be multiple pathways to a specific type of event such as war or alliance formation. This means that other complex bundles of factors—perhaps A, J, and K, or D, J, K, and L—may also be sufficient for E to arise.6

The reality of INJS and multiple causality in the onset of rare event E over time and space means that we cannot simply identify a theory that poses factor A as important and then test it against competing theories that specify factors B or J as important. It may be that A was present at time t through t + 10 but event E only occurred at t + 10. If our investigation shows that the leader of state Y was motivated to initiate E because of A, but that she held off until necessary conditions B and C were in place, then the presence of A and not-E from t to t + 9 does not hurt any theory arguing that A is a critical factor in the cause of E.7 Rather, it is imperative to show by interpreting the documents that the leader of Y was indeed driven by A to initiate E and for the reasons the theory hypothesizes, even if B and C are also important parts of the causal mix that led to E.8

The second claim is that to truly understand why particular event E came about after certain necessary conditions were in place we must establish what causal role the factors A, B, and C played in the arising of E. We need to know more than simply that factors A, B, and C were associated with E. That is, we need to know whether they were propelling leaders toward behaviors that led to E, or whether the factors played more of a facilitating or reinforcing role, or indeed were constraining the leader from taking actions that might otherwise have produced E. These are terms that are often used in academic discourse but rarely defined. A propelling factor is one that directly involves a leader’s ultimate ends and desires or fears—her “reasons” for action. A facilitating factor is one that is incidental to a leader’s ends but needs to be in place before the desired action can be carried out. A constraining factor is something pulling an individual back from doing what she might otherwise want to do. Finally, a reinforcing factor is one that makes the potential effect of a key propelling factor that much more likely to occur.9

When we are investigating historically the complex conjunctural causality underpinning rare events, understanding the functional role played by a variable is critical. Indeed, the very “support” or lack of support of a theory will depend on this understanding. If, for example, a leader’s domestic unpopularity is pushing her to initiate war, but a high level of interdependence is constraining her toward peace in the short term, then the commercial liberal argument would be upheld. Conversely, if the trade environment is actually propelling the leader to choose war to maximize the nation’s security, and the level of domestic popularity is simply a constraining or facilitating factor, then depending on exactly how trade is pushing the leader, the economic realist or trade expectations arguments have potential explanatory force.10 The case studies in Economic Interdependence and War thus seek to establish that combinations of great power interdependence and declining expectations of the future trade environment were critical

---

2 See Copeland 2015, 13, 69f and discussion below.
3 Copeland 2015, 53-69.
4 Copeland 2015, 75-78, esp. fn. 34.
5 Copeland 2015, 76-77; Büthe 2002.
6 See Bennett and Elman 2006; Copeland 2015, 71-72; Mackie 1980; Mahoney and Goertz 2006; Ragin 1987; 2008.
7 Stated differently, because the absence of rare event E—in this book, for example, not-war or peace—is at any point in time overdetermined, the presence of A during a time of not-E does not disprove a connection between A and E, since A is not specified in the theory as a sufficient condition but only as an important necessary condition within an INJS bundle.
8 Copeland 2015, 74f; see also Copeland 2000, 29ff.
9 Copeland 2015, 72f.
10 Copeland 2015, 73.
propelling factors for war, even as the book recognizes that other factors in the INJS mix, including domestic and bureaucratic variables, were often present as facilitating or as reinforcing factors for conflict.

The final methodological claim follows naturally from the first two, namely, that in qualitative research, we should be examining how often a factor was important, in what ways, and under what conditions. The specification of the essential universe of cases within a bounded timeframe helps us to reduce the selection bias that so often plagues qualitative work. It forces us to include in our set of cases ones that might not work well for our pet theory. And by doing so, we can start to identify the conditions under which the theory is indeed most likely to work, or to not work. We might see, for example, that a theory’s causal logic is more likely to be born out in situations of high levels of industrialization and mutual dependence on third parties for key natural resources, and less likely to work when states are simply competing for markets in higher-end luxury goods. By covering a broad range of cases, it might also become clear that a theory’s core factor A is propelling leaders to bring about event E only under conditions M and N, but when conditions B and C are present, factor A typically operates more as a facilitating or constraining factor. Finally, we may find that under certain conditions, factor A from a pet theory sometimes works in conjunction with factors D and J from a competing theory for a specific event E, meaning that both theories answer “part of the puzzle” for why E occurred. Such fine-grained empirical findings should help spur future extensions of a theory, including the specification of the conditions under which a theory’s independent variables will likely be especially salient in a particular INJS mix. They may even lead to the development of new theories that transcend the limitations of the original deductive logic. At the very least, they can encourage an investigation of cases from outside a study’s specified boundaries that will help to hone the understanding of the conditions under which different theories will truly work across space and time.11

All this means that in qualitative testing we should not be seeking definitive tests of whether theories stressing factor A “beat” theories focusing on factors D or J. Rather, knowing that the events we are studying are rare, and thus are only likely to arise when particular bundles of factors are present, we can figure out exactly how often and under what conditions bundles with factor A in them are implicated in the arising of events across time and space, and what specific causal role factor A played in the onset of the events. When there are events E that arise as a result of a combination of A and D, we can then assess both the relative causal salience of A and D in the mix, and whether A and D were both propelling factors, or were doing something else to bring about E. From the perspective of wanting our theories to be practically valuable for policy makers, this approach to research can pay big dividends. It can tell a leader or official when and under what conditions to worry that factor A, or factor D, might lead a nation into a war or a destabilizing crisis. And even if, say, the bundle of D, J, K, and L explains only five percent of the wars over the last two centuries, this fact is still important: if these factors are beginning to manifest, officials can take steps now to reverse the process, knowing that otherwise a war or crisis might break out.

In sum, the methodological set-up of Economic Interdependence and War seeks to offer a balanced approach to the testing of IR theories, one that reduces (but as the symposium shows, does not fully eliminate) some of the problems that have hung over qualitative methodology over the past few decades. By covering the essential universe of cases for a specific time frame, and by examining what functional role variables are playing within an INJS bundle, we can get a better handle on how often a theory works—or does not work—across time and space, and why. Moreover, by examining the cases that a theory cannot explain, we establish a basis for improving the theory and for specifying clearly the conditions under which it is likely to be useful. The study of rare events should always lead us to think in terms of multiple pathways to event E—different complex bundles of factors that can explain the various arisings of E at different times.12 We can still be bold in putting forward variables that we believe drive many, perhaps even the majority, of rare events for a particular time period. But, instead of endlessly arguing over the “master explanations” of phenomenon E across world history, our discussions can focus on the healthier debate about under what set of conditions particular theories should explain outcomes, and, for real-world policy makers, whether current conditions justify the use of one particular theory or another.

References

11 See Copeland 2015, 70f and 94ff.
12 Mackie 1980.
The Empirical Study of Great Power Politics

Timothy McKeown
University of North Carolina, Chapel Hill

If international politics is about "power," then what do governments do once they have it? In Economic Interdependence and War, Dale Copeland suggests that they often use it to gain control of resources on favorable terms, then use that control to leverage additional gains in control and wealth.

Copeland revives a literature that appears to have become moribund—"systemic" theories of international relations. He returns to the venerable liberal versus realist debates and shows that there is new life in old theoretical ideas. By doing so, he moves the theoretical conversation towards a synthesis of these two viewpoints.

His primary tool for accomplishing this is his focus on what he terms "trade expectations," which is used here to refer to the outlook not merely for international trade, but rather a broader array of other economic goods obtained from the rest of the world. When government officials expect that their nation will reap large gains from these international sources, they are inclined to be more peaceful and cooperative. Conversely, when their expectations turn pessimistic they attach a lower value to a peaceful status quo and begin to look for ways to use their available resources and tools of influence to redress their declining fortunes.

While the theory that supports this argument is informal, the basic argument is simple and generally compelling. At its most basic, one could view international politics as countless repeated-play bilateral Prisoners' Dilemma games along the lines of Axelrod. Thus, increasing the payoffs from mutual cooperation dampens the temptation to defect. The larger the mutual gains from cooperation, the more the world resembles the liberal-idealist vision in which creating institutions and practices to realize and lock in joint gains becomes a large part of diplomacy. In a world where those gains dwindle, the chances of aggressive unilateral actions increase, the possibility of war rises, and the international system begins to resemble the tense, balance-of-power world typically depicted by realists.

Economic Interdependence and War engages in an extended theoretical dialogue with idealist and realist theories in the course of making the case for its own trade expectation approach. However, here, the theoretical side of the book will be set aside to focus more deeply on the book's unusual and remarkable empirical strategy. After a chapter that reviews various statistical models of international conflict, the author then sets forth an argument for studying the role of trade expectations by relying completely on case studies, with the balance of book largely devoted to presenting the cases. The justification for proceeding this way rests on several supporting pillars: a critique of statistical approaches, an argument for why case studies can be used to deal with the weaknesses of the statistical approach, a perspective on the nature of expectations and the role of case studies in studying them, and a description of how the case studies will be conducted and used. Each of these elements is discussed below.

The Critique of Statistical Approaches

Copeland is interested in studying events that happen quite infrequently—great power wars, or situations where great power war was a threat but did not actually occur. He identifies only 40 such cases during 1790-1991. He realizes that a rare events logit or probit approach to studying the outbreak of these wars could be taken, but suggests that doing so would not address a deeper problem—the presence of multiple causal pathways to the outcome of war, each of which operates in a highly context-dependent fashion: "It is a self-evident point that rare events in international relations and comparative politics, such as crises, revolutions, and wars, will be a function not of a single complex bundle of [individually necessary, jointly sufficient] conditions but rather different bundles of factors operating with different causal force at different places and times. That is, even when we think in terms of complex conjunctural causality, we must still think in terms of multiple pathways to event E." The combination of complex theory and limited data thus creates an extraordinarily difficult challenge for statistical analysis. Although he does not use this analogy himself, what he is describing seems to be akin to a situation in which a relatively small sample is being mapped into a relatively large, multi-dimensional contingency table, producing an over-abundance of sparsely populated or empty cells. His own discussion stresses the dilemma posed by enriching a statistical model to include interaction terms, suggesting that this complicates the interpretation of statistical results, and that these complications are further deepened when theory implies the incorporation of threshold effects or other non-linearities in the statistical models. While these might be valid concerns at times, their importance depends on the specifics of the model being considered. In any event, it is an issue that is not closely linked to sample size.

Copeland is also skeptical of the general idea that statistical results support the drawing of causal inferences, noting that: "The causal mechanisms that lead to peace or war will be inadequately understood if this is our sole or primary methodology, given that quantitative methods are inherently about correlations and association between variables rather than causality per se." This is a venerable claim: The precise manner by which statistical results inform judgments about causation has been the subject of debate ever since Sewall Wright first

1 Axelrod 1984.
2 Copeland 2015, 77.
3 Copeland 2015, 71-72.
4 Copeland 2015, 51.
presented the method of path analysis and fellow geneticist E.H. Niles criticized his use of it to support claims about causal processes.\(^5\) It is widely agreed that statistical results can inform judgments about hypothesized causal relationships, but that they do not automate the creation of causal theory.\(^6\) The history of these discussions suggests that what is required in making valid causal inferences is less a matter of additional or different evidence than the existence of a causal theory to be tested. Statisticians do not always agree on their precise conception of how statistical work clarifies questions of cause and effect, but they do seem to agree that statistics has a great deal to offer those who are attempting to investigate causation, and that concepts of covariation are close to the heart of most definitions of causality.\(^7\)

Copeland also has a much more specific objection to the use of quantitative analyses in a study that focuses on expectations: He claims that quantitative methods are burdened with an inherent difficulty in dealing with leaders’ expectations because they are forward-looking—expectations are a forecast that affects the current behavior of decision-makers.\(^8\) Just why this poses a problem is not obvious from his discussion. Expectations are indeed forward-looking, but they are formed on the basis of information that has already been acquired. While as a practical matter it might be difficult to write an equation that accurately captures how received information is combined and evaluated to produce a forecast, this is an activity that does not involve information that agents do not already possess.

How Are Case Studies Different?

If we turn from this critique of statistical studies to what researchers do when they write case studies, we encounter a description of case study practice that amounts to “statistics without numbers”—making non-quantitative judgments of description of case study practice that amounts to “statistics when they write case studies, we encounter a description of case study practice that amounts to “statistics nonparametric statistical model of events would also face the previously noted difficulty posed by the relative rarity of cases and the relative richness of the generating causal processes. In this situation, case studies might be the best approach simply because they are the only feasible approach. However, that still leaves open the question of whether the presentation of the cases is best treated as an exercise in the qualitative assessment of covariation. George and McKeown suggested that “process-tracing” methods are an alternative to congruence analysis.\(^11\) That alternative might be especially relevant in a study that focuses on the formation of expectations.

The Nature of Expectations and How to Study Them

If expectations of future economic gain are such a central part of the historical process leading to war or peace, then it matters a great deal how expectations are formed. Copeland is critical of the current literature on this point, claiming that its models are “static and backward-looking.”\(^12\) It is necessary that expectations be backward-looking because they are based on information already received, but they need not be static if information is being updated. Taking note of rational expectations arguments in microeconomics, he posits for himself an even more demanding theoretical assumption about the rationality of decision-makers: not merely that they are rational on average, but that every decision-maker is rational.\(^13\) Disagreements among national decision-makers (it is unclear how he views the possibility of divergent expectations across national boundaries) can occur, but we are not told how they are resolved—only that their individual estimates are assumed to be devoid of systematic bias. Inasmuch as there is no formal model here, it is not clear what is gained by making such a strong assumption, except perhaps providing a justification for not delving into possibilities of divergent perceptions in the case studies. It is easier to see what is lost by such an approach: the opportunity to ground a theory of how expectations affect the

\(^{1}\) Wright 1921; Niles 1922, 1923; Wright 1923.
\(^{2}\) Denis and Legersky 2006.
\(^{3}\) Holland 1986; Glymour 1986.
\(^{4}\) Copeland 2015, 74.
\(^{5}\) George and McKeown 1985.
chances of war in a treatment of them that takes seriously the
cognitive capacities and limits of individuals, the ways in which
their advisory systems operate to provide them with informa-
tion, and the role of larger organizational processes of foreign
ministry bureaucracies, intelligence agencies, and military of-
cers and civilian national security officials, and of leaders’
imformal personal networks and contacts. Case studies of de-
cision processes would seem to be the empirical approach
best suited to uncovering evidence in support of a behavioral
theory of expectations, but that is not how cases are used
here. His empirical chapters have the more modest objective of
establishing just what the decision-makers’ expectations were,
and coming to some judgment about how salient the expecta-
tions were in the decision-makers’ definition of the situation
that they confronted. The cases suggest that, while such ex-
pectations are common, and they frequently seem to be sa-
lent to decision-makers, the role that they play in each case is
heavily context-dependent, and that it is currently not fruitful
to try to develop a single overarching measure of trade expec-
tations. Without having one, we cannot simply re-analyze the
regression models while including a measure of trade expecta-
tions. That might not be such a loss if we can succeed in
establishing that both realist and idealist theories have ne-
glected trade expectations at their peril—a more limited goal,
which seems to have been attained.

Copeland suggests that statistical analyses of causation
merely provide “indirect” evidence of the factors leading to
war,14 and that “in-depth qualitative research has the advan-
tage over quantitative methodologies in that it can unpack the
exact mix of causes that go into particular cases.”15 These claims
seem to suggest that, while quantitative analyses are restricted
to gathering information about variables already postulated as
relevant, case study research can uncover evidence of hereto-
fore unsuspected considerations impinging on governments’
decision processes when they examine government documents
or other sources of evidence about the decision process. If we
agree with Copeland that this is a deficiency, it is one that is
not inherent in statistical analysis, as one can always expand
the data-set and re-analyze the data if one has happened upon
new information deemed relevant. But the extant quantitative
literature is only “indirect” because of the kinds of variables
that quantitative research has generally considered, and which
are featured in his discussion of quantitative research in chap-
ter two: they are publicly observable characteristics of gov-
ernments or nations, not variables chosen to capture the na-
ture of the information flow reaching high-level decision mak-
ers. As such, the quantitative models implicitly assume that
these publicly observable national features affect all decision-
makers in a uniform way—differences from time to time or gov-
ernment to government in how information is processed are
assumed to be theoretically unimportant, though from an em-
pirical standpoint their possible existence might motivate a
fixed-effects approach in a pooled model. Copeland’s assess-
ment suggests that he realizes that something is lost when the
decision process is not systematically examined, but the fail-
ure to do so is not due to the choice of research method. For
example, a text-as-data researcher could upload a large corpus
of foreign policy documents and analyze them statistically,
and even evaluate how changes in the content of the corpus
are related to changes in policy decisions. This evidence would
be a good deal more “direct” in the sense that it is taken from
internal decision processes, but what has changed is less the
reliance on statistical method than the model used and the
choice of variables to be included.

Using government documents and other sources to trace
the evolution of the decision process in a case could provide
insight into the inner thinking and planning of leaders. This
seems to be Copeland’s approach, and he claims in his theory
chapter that “the best way to investigate [expectations] is
through careful documentary analysis.”16 This is what he does
for the cases where the supply of primary documents is rela-
tively abundant. When it is not, he relies on historian’s ac-
counts. Because these are typically based on a reading of
primary documents, they provide evidence about the decision
process, but with an intervening layer of analysis. Since the
methods used by individual historians to support their judg-
ments about the decision process are typically not recover-
able—they seldom have “methods” chapters in their books—
we are left, as Copeland is, with the choice of either ignoring
their conclusions, or else treating them as a panel of experts,
each of whom is highly knowledgeable about the details of a
case, but who is not generally in a position to help us recon-
struct the research process that led to their conclusions. In a
scholarly age that is highly concerned with the transparency
and reproducibility of the research process, this seems like a
significant limitation, but another feature of the current era—
the exploitation of “crowdsourced” judgments and actions in
support of research—provides a way to think about how a
large body of historical writings could be pressed into service
to support research on international relations even when the
basis for any single judgment is unclear. However, if any re-
search strategy deserves to be described as “indirect”, this
one certainly does.

How should we assess the threats to the validity of causal
inferences drawn from archival material (or, in some cases, from
others’ readings of the archival material)? Copeland does dis-
cuss the possibility that one’s own theoretical framework fil-
ters the reading of the case material so as to exclude discord-
ant information. “The only thing one can do, therefore, is to
seek to be as self-conscious as possible about such potential
biases.”17 Although he adds that we need “objective handling
of documentary evidence” and a “careful methodological
setup,”18 some more specifics about these points would be
welcome. Trying to be honest is never a bad idea, but it is not
even: document creation, distribution, redaction or destruc-
tion, and public release can all be viewed as the product of

16 Copeland 2015, 74.
17 Copeland 2015, 95.
18 Copeland 2015, 96.
strategic choices, and hence it is wise for scholars to adopt their own strategies for coping with the challenges posed by strategic release and denial. The preserved and released written record of foreign policy decisions over-emphasizes “national interest” justifications for policy choices, while de-emphasizing the influence of informal communications, especially with private actors (in the US case, *Foreign Relations of the United States*, the official compendium of declassified material, rarely includes any documentation of private communication with US government officials over foreign policy). It also tends to slight the importance of covert operations, which, along with signals intelligence and code-breaking, are typically among the most closely held material and are often released very late or not at all.

**Conclusion**

Copeland’s book opens the door to three significant accomplishments in the development of a deeper understanding of international relations. First, it clears a path forward towards a realist-idealist theoretical synthesis by identifying a set of conditions that could account for why each of these two general orientations successfully describes international politics only part of the time. Second, it re-opens discussion of a topic that was at the forefront of theorizing about international relations in the period before the Cold War and the intellectual dominance of balance of power thinking—the relation of “power” to “plenty.” Paul Reinsch, often regarded as the founder of modern American scholarship on international relations, opened his discussion of world politics at the dawn of the twentieth century not with a discussion of naval competition or European alliance systems, but with a map of the Chinese railway system. Jacob Viner inaugurated the journal *World Politics* with an extended discussion of how scholars of the previous century had struggled to make sense of the relation between the pursuit of wealth and the pursuit of politico-military advantage. Albert Hirschman developed a highly influential theory of how Nazi Germany used its trade policy not simply to acquire wealth, but as an instrument of national power. These lines of inquiry receded with the onset of the Cold War, but not because they had been resolved. Third, his account directs our attention to the possibility and desirability of writing a behavioral theory of foreign policy-making that is richer and more accurate than those based on a treatment that posits unitary rational actors. This is an exciting set of research objectives to pursue, and we can thank Dale Copeland for lighting the way forward.

---

19 Colaresi 2014.
20 Anderson 1981.
21 Reinsch 1925 [1902].
22 Viner 1948.
23 Hirschman 1945.

**References**

Evaluating “Competing” Explanations in Economic Interdependence and War

Sherry Zaks
Dartmouth College

Dale C. Copeland’s Economic Interdependence and War tackles head on one of the central debates in grand IR theory: whether and how economic interdependence between great powers affects the likelihood of war. With realist scholars arguing that interdependence is a source of vulnerability and a cause of conflict, and liberal scholars arguing that interdependence is a source of stability and a mitigator of conflict, IR’s main paradigms remain unable to fully account for both the massive amount of interdependence among great powers and the conflicts between them. Initially framing his book as a sort of middle ground, Copeland argues that the “impact [of interdependence] can cut both ways” and so we must shift the question to “when and under what conditions will the trade [... ] between nations lead to either peace or military conflict?”

Copeland argues that the bifurcating factor is whether leaders have positive or negative expectations about their country’s future trade relationships: anticipation of continuity or prosperity should foster restraint, whereas anticipation of decline or cut-offs should prompt aggression.

The theoretical importance of this variable is nontrivial and Copeland successfully demonstrates through rich archival analysis that leaders’ trade expectations often play a central role in conflict initiation and restraint. However, the theoretical framing and corresponding analysis suffer from two related shortcomings that leave many questions unanswered and ultimately detract from the potential contribution of this work. First, Copeland’s attempt to situate trade expectations theory vis-à-vis realism and liberalism is at best ambiguous. Copeland is unclear about whether his framework is working against, alongside, or as a part of the alternative explanations he examines. This form of theoretical ambiguity has direct methodological implications. As such, the second shortcoming is a research design that is incapable of fully testing the range of predictions that follow from the respective frameworks and the relationships among them. I draw on the Relationships Among Rivals framework—a set of analytic guidelines for testing nested and competing hypotheses—to examine the theoretical and methodological weaknesses and to offer suggestions for future work examining the role of trade expectations.

Trade Expectations Theory vis-à-vis Realism and Liberalism

Beginning with our training and evident up through our publication standards, political science exhibits a strong preference—if not bias—in favor of novel theoretical frameworks that quash competing explanations. While this preference is understandable in terms of seeking greater contributions, it often manifests as an ambiguity in theoretical framing and an incompleteness in empirical analyses as authors erroneously frame their theories as occupying a mutually exclusive position with respect to alternative explanations. In reality—and much more frequently—two theories may also be coincident (i.e., both could simultaneously account for the phenomenon, with evidence in favor of one not affecting the other), congruent (i.e., not only do both account for the outcome, but evidence in favor of one theory also supports the other theory), or inclusive (i.e., one theory represents a novel extension of the other).

Without specifying how alternative explanations relate to one another, scholars risk dismissing alternative theories without sufficient justification and leaving readers with dubious guidance on how and when to use the different frameworks moving forward.

Unfortunately, Copeland is not immune from this criticism. At various times throughout the book, Copeland presents the three main theories he addresses as mutually exclusive explanations, such as when he writes:

“[t]he main goal of the case studies is to test the logic of trade expectations theory directly against its main alternatives” (emphasis added), and

“[t]he empirical chapters have shown...the superior explanatory power of trade expectations theory over its rivals.”

At other times, he frames them as coincident explanations, in which multiple logics contribute to the outcome:

“rare events in international relations [...] will be a function of [...] different bundles of factors operating with different causal force at different places and times,” or as congruent, in which evidence for one theory also supports the alternatives:

“[B]oth liberals and realists have a point. They have just not specified the conditions under which they might be right. [Trade expectations theory] will help resolve this problem,”

“[t]rade expectations theory [...] offers a new variable [...] as a way to link liberal theory’s emphasis on the benefits from trade with realism’s concern for the potentially significant costs of adjustment that a state would face were it to be cut off...”

At yet other times, Copeland portrays the relationship between the three theories as inclusive, in that one theory constitutes a

---

1 Copeland 2015, 1.
2 Zaks 2017.
3 Zaks 2017, 348.
4 Copeland 2015, 50.
5 Copeland 2015, 432.
6 Copeland 2015, 77.
7 Copeland 2015, 24.
8 Copeland 2015, 429.
fully subsumed extension of another:

“At its foundation, [trade expectations theory] is fundamentally realist.”

In this section, I discuss the source and consequences of this ambiguity, identify the actual relationships between trade expectations theory and its “competitors,” and lay out the implications for research design and future work in international relations theory.

Having a clear sense of where trade expectations theory sits relative to alternative explanations of conflict under interdependence is especially important in this context. This is the case because Copeland does not merely test his variable against some other variables. Rather, he situates trade expectations theory to take on the two main paradigms in international relations in an attempt to explain one of the most high-stakes outcomes. Any support for Copeland’s framework will necessarily have critical implications for the future of international relations theory. As a result, this analysis demands explicit guidance on how (and whether) to apply realism, liberalism, and/or trade expectations theory moving forward. On this front, however, the reader is left wanting.

This ambiguity is primarily attributable to Copeland’s underlying ideological preference for realist theory. He argues that, by maintaining the foundational assumption that states are primarily concerned with security (rather than material gains), trade expectations theory is “fundamentally realist” in nature. Upon closer consideration, however, this framework need not be confined to either paradigm—arguably, one could make an equally compelling case that trade expectations theory is fundamentally a liberal one. To be sure, where a new hypothesis can help bolster any theoretical framework, researchers should be explicit about how it fits into an existing theory, what needs to change, and how to go about testing it. Maintaining paradigmatic preferences can—and in this case does—undermine the contribution of new theories.

Drawing on the framework for specifying the relationships among rival hypotheses outlined in Zaks, I demonstrate below that the trade expectations variable represents a useful extension to both economic realism and liberalism that should be applied in contexts of great power dependence in order to make more robust predictions about the likelihood and timing of conflict. Copeland’s framework is not intended as a replacement for either paradigm, nor does support for trade expectations theory directly undermine its competitors. Moreover, the logic of trade expectations is largely consistent with both paradigms and the predictions on the outcome differ only where realism and liberalism fall short. As such, ruling out mutual exclusivity is straightforward. The question now becomes whether trade expectations is a variable that exists independently of realist and liberal logic, or whether this framework can be incorporated into both of the paradigms. I address each one of these in turn.

Turning first to realism, I argue that, in its current state, realism does not offer a unique hypothesis that accounts for the variability in the likelihood of conflict under interdependence. Therefore, it does not make much sense to think about realism and trade expectations theory as simultaneously and independently offering an explanation for the outcome. Rather, realism posits that the vulnerability resulting from dependence should drive states toward more aggressive tactics. Trade expectations theory incorporates this prediction, but adds nuance to it by considering both the benefits of trade and an assessment of the future. As such, the trade expectations hypothesis is both fully compatible with the realist paradigm and picks up the slack where realist predictions fail. Trade expectations theory therefore constitutes an extension that enhances the predictive power of the realist paradigm under conditions of great power interdependence.

Liberalism, in contrast, offers a set of specific hypotheses about the conditions under which conflict is more or less likely in contexts of great power interdependence. Domestic-level variables (such as regime type or economy type), commitment problems, and the “shadow of the future” have all been advanced by liberal scholars to account for interstate conflict and peace. It is worth noting that Copeland does not disaggregate liberal theory into those constituent hypotheses. I argue, however, that doing so is crucial to achieving precision in both theoretical and methodological discussions. Trade expectations theory is certainly not exclusive to the first two hypotheses—i.e., finding support for a leader’s fear of future cut-offs would not make it impossible for regime type or bargaining problems to also play a comparable role. Both factors could easily coexist, but independently of trade expectations. Thus, Copeland’s framework is coincident with these two liberal explanations: any or all of the three can be contributing factors to a given instance of conflict or peace and each must be tested separately.

As for the “shadow of the future” hypothesis—the liberal notion that two states’ anticipation of multiple future interactions helps promote cooperation and mitigate conflict between them—trade expectations theory represents a congruent and inclusive extension to liberalism on this dimension. In essence (though he does not frame it as such) Copeland’s framework posits that the shadow of the future is a variable rather than a constant: leaders do look to the future of their trade relationships to inform present behavior, but that future is not always assessed favorably. Moreover, trade expectations theory draws on other liberal notions to account for variation, such as changes in leadership or regime. Thus, trade expectations theory can be framed as a modification and enhancement of the shadow of the future hypothesis, which means it can also be

---

9 Copeland 2015, 27.
10 Copeland 2015, 27.
11 Specifically, Copeland acknowledges that dependence is in part defined by the benefits of trade and that there are circumstances in which the benefits of trade outweigh vulnerability (35). He also departs from structural realist assumptions by acknowledging the role of unit-level characteristics in one’s trade partner (39, 41).
13 Copeland 2015, 22.
14 Fearon 1998.
fully integrated into the liberal paradigm by incorporating additional considerations (such as threat and security) into predictions based on future trade assessments.

In sum, by constituting a compatible extension of both realism and liberalism, the trade expectations hypotheses represent an unprecedented contribution to international relations theory. The trade expectations framework borrows key insights from competing paradigms in order to better parse the effect of interdependence on the likelihood of conflict. From a theoretical standpoint, Copeland implicitly demonstrates that both paradigms are excessively blunt, but in different directions: realism is naively pessimistic about the effects of interdependence and liberalism is naively optimistic on the same score. Then, he manages to develop a single theoretical framework capable of adding explanatory and predictive accuracy to both at once. Unfortunately, Copeland’s predisposition toward realism and ambiguous framing of where trade expectations theory sits with respect to alternative accounts of conflict overshadows the true scale of this contribution. In reality, scholars from both traditions can proceed by incorporating the insights from trade expectations theory into either lens.

Methodological Implications

This more nuanced identification of how trade expectations theory relates to alternative explanations of conflict under interdependence has critical implications for research design and inference. I now seek to show that the nature of the relationships among competing hypotheses shapes the type of tests one can conduct as well as the inferences one can draw.

Unfortunately, Copeland’s empirical analysis falls short on two fronts. First, his analysis and the implications are obfuscated by the tendency to force tests of multiple theories to conform to the “testing Theory X against Theory Y” construction. In reality, this purported goal of Copeland’s empirical tests is only possible under a limited number of circumstances and is constrained directly by the relationships among competing theories. Second, drawing on the central tenets of each theoretical framework, I argue that his identification and selection of cases is too narrow to address the full range of hypotheses and relationships he wishes to test.

Constructing Tests

Copeland notes that the goal of his empirical analysis is not only to establish the overall importance of trade expectations, but to identify how strong of a role they played relative to the “bundles” of other factors present in each case. Yet, one of the primary insights that follows from the characterization of relationships among the “competing” hypotheses above is that testing the relative “causal force” is only possible when two explanations can simultaneously but independently bring about the outcome (i.e., under relationships of coincidence). Under mutual exclusivity, we would not consider asking about the relative force, since only one explanation is possible at a time. Instead, researchers who want to get some leverage on this question are limited to testing 1) which explanation is present in a given case, and 2) which theory explains more outcomes across the universe of cases (i.e., the frequency with which one or the other has explanatory power). Under inclusive relationships (i.e., when Theory Y is integrated into Theory X as an extension), the question of relative salience is non-sensical: if one is extending an existing theory, we would not consider the two theories to be separate. Researchers are limited to testing whether the expanded theory 1) performs better (either in terms of prediction or explanation) in a given case, and 2) whether it generates more accurate predictions than it did prior to the extension.

Thus, to get traction on the analytic purchase of trade expectations theory vis-à-vis existing explanations, Copeland must construct tests and select cases that are tailored to the constraints described above. Testing trade expectations theory “against” realism is a matter of constructing tests to identify whether incorporating this variable enhances realism’s predictive (and explanatory) capacity. Given that trade expectations theory is best situated as a modification and extension of realism, the tests should be constructed to assess whether this expanded version of realism predicts and explains more cases with better precision than it did in its original state.

The depth and breadth of evidence Copeland analyzes to assess leaders’ evaluations of future trade relations is impressive by any measure. However, the analysis becomes clunky due to the unclear implications of supporting one theory over another. By forcing himself into the “testing theory X against theory Y” box, Copeland overlooks and/or undersells his own contribution to both realist and liberal theories. In some cases, for example, he concludes that “trade expectations theory and economic realism [are] best,” while in other instances, he argues that his framework outperforms realism. These conclusions, however, leave two questions unanswered that harken directly back to the relationship between his framework and realism. First, how are researchers supposed to proceed—should they be employing both theories or different ones under different conditions? Second, how can we reconcile that trade expectations theory can simultaneously outperform realism yet also be a “fundamentally realist” theory? Thus, while his tests demonstrate in some sense that the current state of economic realism is underpowered and that trade expectations theory performs better, his research design is incapable of testing the extent to which the trade expectations hypothesis enhances the predictive power of realism in the context of interdependence. As a result, his readers may likely come away feeling as though they learned something, but uncertain of how to apply it in a systematic way as they move forward.

Testing trade expectations theory with respect to liberalism is necessarily a more complex task. Since liberalism poses a variety of explanations intended to parse conflict and peace under interdependence, he must proceed carefully with case selection, evidence gathering, and analysis to test the full range of variables and predictions. Copeland’s assessment of liberalism highlights additional problems with his analysis. First, as

---

15 Copeland 2015, 77.
16 Copeland 2015, 88 (in Table 2.7)
17 Copeland 2015, 27.
I alluded to previously, Copeland does not fully disaggregate liberalism into its three constituent hypotheses regarding how conflict (and peace) can result from interdependence. Since the first two liberal explanations—the role of unit-level variables and bargaining dynamics—are coincident with his theory, each should have been addressed by searching for confirming and disconfirming evidence for each hypothesis independently.

The second issue with testing trade expectations against liberalism is that Copeland at no point acknowledges the shadow of the future hypothesis, which is both surprising and problematic given his steadfast calls for a more dynamic and forward-looking theory of state interactions. Because liberalism includes a dynamic theory, it is all the more important for Copeland to explicitly address how trade expectations theory differs—whether it should be incorporated as a modification or replacement, and what the implications are for the shadow of the future concept moving forward. By not disaggregating liberalism into its constituent theories, the reader is often left with a surface-level rejection of the liberal paradigm, but with no specific identification of where liberalism goes wrong.

**An Incomplete Universe: Case Identification and Selection**

The central tenets of each theory as well as the relationships among them should be synthesized to inform how cases are defined and selected for a given analysis. Datasets ought to be constructed according to theoretically-informed inclusion criteria to guard against selection effects and confirmation bias. I argue that the universe of cases Copeland identifies is incomplete in a way that systematically favors trade expectation and realist explanations (yet still prevents even the fullest testing of these hypotheses). I do feel it is important to begin this discussion by giving Copeland full marks on the ambition of this analysis. Many qualitative analyses include one tenth the number of case studies—or fewer—found in Copeland’s book and are still considered thorough. As such, the following critique is not to argue in favor of increasing the N for the N’s sake, but to argue that the universe is systematically incomplete: the cases it omits are likely to reveal something different about the way the world works.

The problem manifests first by defining “case periods” as necessarily comprising “great power cases of crisis and war” and requiring that they be “marked by a particularly salient issue or important event.” By restricting the sampling frame to interactions that eventually culminated in a crisis, Copeland, despite his argument to the contrary, is selecting out cases of long-term stability. The justification he puts forth—that by considering the non-conflictual periods surrounding the crisis, he does address negative outcomes on the dependent variable—only goes so far to mitigate this problem. While this tactic demonstrates that trade expectations theory outperforms previous explanations in terms of predicting the specific timing of a given conflict, it still only examines instances in which crisis was imminent. Yet, for every great power conflict that arose in the context of economic interdependence, one can imagine that there are many other interdependent dyads where no conflict ever occurred.

The omission of non-conflict interactions is problematic on several dimensions. First, it narrows the scope of cases in which Copeland could and should be testing the predictive value of trade expectations theory. According to the framework, when states have positive assessments of future stability in their trade relationships they should be less likely to engage in conflict. Consequently, non-conflict cases are just as important to evaluating trade expectations theory as the great power crises are. Furthermore, these cases are especially critical to evaluating how trade expectations theory performs relative to liberalism. Since the foundational assumptions in liberalism lead scholars to predict that on the whole, economic interdependence is a mitigator, rather than a facilitator, of conflict, rejecting the liberal paradigm in favor of trade expectations theory is nearly impossible without engaging liberal theory on its own turf. This issue is the qualitative incarnation of the nonignorability problem addressed in King: the missing data is in part a function of the explanatory and outcome variables. Specifically, the outcome No Conflict—which is a crucial part of the theoretical frameworks under examination—is highly correlated with missingness in Copeland’s dataset.

**Conclusion**

*Economic Interdependence and War* is a behemoth of empirical work on two centuries of great power conflicts. Moreover, Copeland successfully demonstrates that leaders’ assessments of the future stability of their trade relationships play a central role in many of these crises. Unfortunately, a critical theoretical oversight and the corresponding methodological implications undermine the extent of the book’s contribution. As a result, readers are likely to come away certain they have learned something, but uncertain of precisely how to synthesize their new insight with their existing knowledge.

In this review, I have highlighted the importance of identifying the relationships among competing hypotheses and breaking out of the ingrained frame of the mutual exclusivity. Indeed, Copeland himself is clearly aware that in many cases, multiple explanations may work at once to bring about an outcome. But, by lacking a comprehensive framework for identifying the range of ways hypotheses may relate to one another and the implications for research design and analysis, even the most careful scholars can fall back onto the crutch of “testing Theory X against Theory Y.” Identifying the relationships among competing explanations informs with greater specificity the types of tests that one can conduct and, in turn, should inform how the universe of cases is defined and how specific cases are selected for analysis.

When Copeland presents trade expectations theory in a theoretical context, his characterization of the relationship be-

---

18 I.e., finding support for trade expectations theory has no effect on the validity of these other hypotheses.

19 Copeland 2015, 79.

20 Copeland 2015, 77.

21 King 2001, 51.
between his framework and its predecessors leans more toward coincidence or congruence (i.e., that the explanations may work together or that one may be an extension of the other). Yet, when he moves on to discuss his research design, the trade expectations hypothesis is suddenly framed as occupying an exclusive position. This level of ambiguity about where trade expectations theory sits relative to realism and liberalism cascades through the analysis and has both theoretical and methodological implications. On the theoretical side, this framing leaves the reader wondering what it is we are going to get from the book: is it an extension to realism or liberalism? Is it a fusion of both? Is it a replacement? On the methodological side, readers are left wondering how, for example, to test a theory that is both “competing with” realism and “fundamentally realist.”

Copeland concludes by positing that “trade expectations theory resolves the problems for established liberal and realist theories.”22 In reality, however, his analysis falls short of this goal. Despite providing an impressive wealth of evidence in favor of the trade expectations hypothesis, the theoretical and methodological ambiguities muddy the inferences we can draw and the implications for future research. The crucial addendum is that the trade expectations theory has the latent potential to achieve this resolution between the realist and liberal paradigms and to guide future applications of both. By recasting trade expectations as an extension of both theories, Copeland could demonstrate precisely the need to modify both main IR paradigms to more accurately deal with conflict under interdependence.

References


22 Copeland 2015, 428.
books with the varieties of liberal peace theory. Whatever the details, however, the basic form is one of making war relatively expensive (or peace relatively cheap). Commercial loses associated with conflict pose costs for leaders and countries (economists call these opportunity costs) that, if anticipated, can lead them to prefer peace more often. It is important to note that in no form does this argument fail to put itself in the future. If these costs, or the opportunities to avoid them, are in the past then subsequent action can do no good (or cause no more harm). It is only when one’s actions today shape outcomes tomorrow that opportunity costs matter for decision making. Similarly, a failure to anticipate opportunity costs makes them irrelevant.

In Copeland’s conception, as in other opportunity cost arguments, nations still want to fight. However, the text argues for a key theoretical innovation in focusing on future economic gains or losses. But the distinction between future and past trade that Copeland emphasizes is not salient. Specifically, whether states are currently not trading, but will if they remain at peace, or whether states are currently trading and will continue to do so if they do not fight, is said to prove pivotal in Economic Interdependence and War. But there is nothing in the opportunity cost framework, or in Copeland’s particular version of the theory, that would make distinguishing among different pasts germane. Both are functionally equivalent. In each case, it is anticipation of the future that drives choices.

In basic analytical terms, the effect of trade in the opportunity cost framework is like that of altering the payoffs in the famous Prisoners’ dilemma game to transform it into the almost equally famous chicken game. If the cost of war goes up enough, both sides prefer conceding to mutual destruction. Of course, each side still prefers an outcome where it is the adversary that swerves or backs away. A similar challenge exists between trade partners. One need to look no further than the current headlines to see two states, the United States and China, trying to cow one another over the imposition of risky actions made possible not in spite of, but because of, trade.

This does not preclude Copeland’s conception of the mechanism underlying trade’s effect on peace from operating effectively. However, at least three problems with this logic remain. First, how much is enough? Big wars are very expensive. It is a rare case where two nations’ bilateral trade is on a scale that would make eliminating all trade (let alone some of it) sufficiently costly to fundamentally alter the accounting price of a war, or even of a significant militarized dispute, once states have reached the point where they consider military escalation of a dispute a rational choice. In my own research, I have found that in the post-World War II era, less than four percent of all dyads look this way.

Second, there is the vexing question of endogeneity, here in the sense of reverse causation. Endogeneity implies that the states that are most likely to fight are least likely to trade together. For example, the United States worked hard to prevent U.S. and other western merchants from trading with the Soviet Union. Copeland dismisses these and other endogeneity problems rather blithely. It is not clear why, since they are of considerable concern in the literature. Third and most vexing, I do not think Copeland’s theory succeeds at reconciling contrasting predictions of conflict causation used by different paradigms. The admittedly implicit theory of war underlying liberalism emphasizes a cost-benefit calculus. Realist theories rely on some form of power relations. Realism does not intensively consider war costs (though these appear in the nuclear domain). So, how does one reconcile the predicted effects borne of opportunity costs with relative power, in an environment where prosperity itself may prove to be a problem? Copeland’s solution is ingenious but confused, and confusing. The simplest description is that, in prospect, opportunity costs somehow manage to intersect with and remedy concerns about relative power. If a state expects future trade to make it more powerful, it will not fight today. If instead states perceive that they will be prevented from prospering, they can prefer war. “Such a state will tend to believe that without access to […] markets needed for its economic health, its economy will start to fall relative to other less vulnerable actors.”

This sounds very plausible, especially if one’s mind is still focused on the liberal logic of war, as well as peace. But there are at least two problems in attempting to fuse liberal (costs) and Realist (power) conceptions in this way. First, let us assume for a moment that everyone can see the future equally, so that the future is subject to common conjecture (everyone can see and evaluate everyone else’s problem). If I know that by continuing to trade instead of fight I will be stronger in the future relative to you, then of course this is terrific for me. But Realist you is not going to like this future. If instead you will grow more from trading than I, then Realist me hates interdependence and you are the happy one. Looking into the future, if we agree on what we see, we cannot escape the Realist relative gains problem.

The second problem with using the future to segue from a focus on costs to power has to do with perceptions. My initial supposition about what leaders know about the future must be relaxed. Whether it is world affairs or labor relations, the origins of disputes seem most often to lie in what actors do not know, rather than what they do. In fact, the text is clear that leaders do not know the future. One of the key features of being forward-looking is that one can get things wrong—leaders may not have the same beliefs about their future economic prospects. If there is no common conjecture, then war in Copeland’s world depends on whether Realist leaders anticipate compatible or incompatible effects of interdependence. If both you and I think that we will be relatively advantaged by trade, then we can remain at peace, at least for a while. If instead either of us is pessimistic about the magnitude of future

---

1 For examples of such books, see Doyle’s Ways of War and Peace and Russett’s Grasping the Democratic Peace.
2 Snyder 1961.
3 Gartzke Typescript.
4 See, for example, Morrow’s “How Could Trade Affect Conflict?”
5 Copeland 2015, 2.
economic gains, then we may prefer war. This of course means that the theory really hinges on the question of beliefs and their origins, since what leaders happen to anticipate determines how they will behave, determining whether trade begets peace or war.

There is an enormous literature on decision making and human cognition. It is unreasonable to expect an author to capture all or even most of what other authors are writing. But given the centrality of perception to assessing the accuracy of predictions from Copeland’s framework, some additional detail on how leaders form beliefs would be especially helpful. Here, the book says too little.

Material variables condition who gets what in this story, while the information actors have about material variables condition how they go about getting there. Copeland ends up with this conclusion without consciously embracing it, discussing the process of discovery about relative power and benefits in a manner not unlike Bayes Rule, but without explicitly addressing key insights that follow from tying uncertainty to decision-making and conflict in this manner.

The result is no longer a story about opportunity costs. It is instead a story about beliefs about opportunity costs, or indeed beliefs in general. If states have incompatible conceptions of the effects of trade, whether real or imagined, this can lead them to fight even when the actual mutual benefits of trade should have prevented conflict. Copeland notes this effect without tying it to the appropriate causal mechanism of conflict. If what leaders believe about trade is at the core of whether they fight, then it is beliefs that drive conflict, not material conditions.

In sum, Copeland’s book seeks to meld liberal commercial peace theory with a Realist account of the origins of great power war. These two theories have fundamentally different explanations for the origins of conflict and thus are difficult to reconcile. Copeland’s solution is to look to the future, focusing attention on what leaders anticipate about events and circumstances. However, this does not actually resolve logical tensions between the two paradigms. Nor does it respond effectively to fundamental criticisms of each perspective’s theory of war. It also leads to important questions about the origins of perceptions that are not resolved in the text. These structural problems are obscured under an enormous amount of description and argument.

The Methodological Approach of Economic Interdependence and War

I thus turn, briefly, to addressing the extremely detailed empirical research in the text. There is much in Copeland’s extensive case studies that informs and provokes, making it impossible to do it justice in a short essay. So I will just address a few key issues.

A basic problem, which Copeland disputes, involves endogeneity. The text concentrates on great powers. Separating states into big and small appears to be a hold-over from the billiard ball days, and from an age when anything more than a dichotomy overloaded the human imagination. But focusing on great powers is likely to greatly worsen the endogeneity problem discussed above: One of the factors determining major power status is economic size, and economic size is increasingly driven by integration into the global economy. Thus, Copeland’s sampling procedure incorporates his key independent variable.

A related concern about Copeland’s case selection, as comprehensive as the empirical analysis might appear to be, is that it inhibits our ability to assess alternative explanations. One way in which war can be averted, possibly helped by trade, is if actors lack the basis for competition or conflict in the first place. This is positive (rather than negative) peace. This is how the West was tamed, not by guns but by civilization. In settings as diverse as marriage, the workplace, and the European Union, constituent actors have found sufficient common cause to interact cooperatively. Institutions and norms bridge gaps not to compel, but to facilitate cooperation. Workers in knowledge industries cannot be effectively forced to be creative. Instead, they are given incentives to further the interest of their employers by sharing a stake in the firm’s future. Conquest in such an environment is counterproductive, since there is no practical way to ensure continued productivity and profit except by allowing workers to remain unfettered by coercion, which undermines conquest. In the second half of the 20th Century, the United States as hegemon has not conquered to rule, possess and extract. Instead, it has led a coalition of nations, made more powerful and willing based on prosperity and collective security. Common interests have made war in the Western world all but unthinkable for over seven decades. Copeland recognizes this important peace-producing dynamic, but he instead focuses on explaining the tensions between West and East.

Another issue has to do with inference in an inherently probabilistic setting. The assertion in modern analysis is not that states that do not trade always fight or that those that trade never do. Copeland is mindful of this and too sophisticated to suggest that interdependence is more than one among many factors influencing international affairs. Still, this poses two challenges for his chosen mode of analysis. First, the marginal impact of a variable or process depends on how often it makes a difference in terms of given outcomes in a population of actors or circumstances. How do we know whether the cases Copeland has chosen to study are representative of tendencies in world affairs? How can we tell how much trade matters if the examples he shows us are not representative? One indication of a clear problem in case selection is his focus on rivals; one cannot expect to learn a great deal about peace by looking disproportionately at those that fail to achieve it. Trade may operate at an earlier stage, preventing states from becoming rivals. Second, case studies, even a considerable number of them, cannot clearly delineate tendencies.

There are also a number of admittedly technical and specialized errors that will mar the text in the eyes of experts, thought...
they are unlikely to be noted by general readers. Copeland claims to have replicated statistical studies by other authors. But in doing so, he makes glaring omissions. For example, he incorrectly interprets interaction terms from tabular results and uses models to make out-of-sample predictions.

Another concern about the analytical methods is that the text notes the importance of both exogenous and endogenous factors in a leader’s choice of interdependence. This implies a two-stage process and creates challenges for inference (even under the best of conditions). For example, it is not clear how the text is able to resolve whether causes are endogenous or exogenous since this depends in part on a prior stage of causation. If domestic politics is treated as an exogenous factor, for example, but the choice of regime type is influenced by trade and other variables, then regime type is endogenous. Copeland identifies at least six exogenous factors as confounding variables in the analysis. The variables should be factored systematically into case selection, as well as explication within cases. Yet, this is difficult, given limited degrees of freedom (in other words, there are more variables than cases).

Finally, Copeland’s defense of his methodological approach is that he is analyzing rare events. The discussion of the strengths and weaknesses of quantitative versus qualitative methods for assessing rare events in a multi-causal environment highlights basic challenges to inference facing all methods involving observational data. Over-sampling on failure (in this case war) does not necessarily tell us more about the causes of peace, especially if one believes that causation is multiple and jointly contingent. And there is lots of peace in the world, for a multitude of reasons, so to the extent that Copeland in fact has an explanation for peace, he is, fortunately, not studying a rare event at all.

Ignoring what one wants to explain because the other thing is rare gets the whole enterprise backward. Tests of theories of commercial peace involve contrasting interdependence with many factors that interfere with conflict, such as distance, interests, capabilities, alliances, etc. By looking mostly just at failures, one cannot distinguish the causes of the overwhelming proportion of successes. Again, because the discussion is so detailed, the words sound right, and many readers have no basis on which to doubt the veracity of the author’s claims, highly uncertain inferences may be believed as assertions of fact.

**Conclusion**

Interdependence as a setting for research is challenging (some might say daunting), precisely because the processes involved are multifaceted and complex. Copeland’s book is ambitious, and in the best sense of that word it moves research forward by encapsulating both insights and areas deserving additional attention. I have focused disproportionately on areas where I think it falls short, but this is only because the goal of promoting peace and international stability through a better understanding of commerce and conflict is clearly one of the most important in the modern world. I applaud the aspiration of the author to make the world better, more peaceful. At the same time, I fear Copeland’s particular approach, seeking to synthesize two traditions that each have problems addressing the fundamental process of interest, and conflict itself, is not likely to prove productive, at least partly due to the methodological choices he makes. I hope that future iterations of this scholarly objective are even more careful in their causal logic and in the methods they use for the empirical assessment of that logic in our fractious world.

**References**

Rare Events and Mixed-Methods Research: Shaping the Agenda for the Future

Dale C. Copeland
University of Virginia


I very much appreciate the opportunity to respond to three thoughtful critiques of the methodology of my book, Economic Interdependence and War.1 Timothy McKeown, Sherry Zaks, and Erik Gartzke offer important and constructive comments on the mixed-methods approach I adopted in the book. Adopting the positive spirit found in the critiques, I will not attempt a blow-by-blow rebuttal of their arguments, but rather seek to show how some of their insights can be used to advance the book’s original intention—that of building a distinctive approach to mixed-methods research for the study of rare events in international relations (IR) and comparative politics (CP). The big take-away point from this essay is that there is no one best way to do social science research, even in a perfect situation where we have a great deal of evidence available for both quantitative and qualitative analyses. The right overall approach, and indeed the right balance or “mix” of quantitative and qualitative methods, depends fundamentally on the type of phenomenon one is trying to explain and the type of relationship between independent and dependent variables that one is considering. Once we get a handle on these elements, I will argue, many of the issues and concerns broached by McKeown, Zaks, Gartzke and other scholars about the role of qualitative research in the broader project of social science can be addressed and at least partially resolved.

Two questions animate the following discussion. First, to what extent does the type of event that a researcher wishes to explain affect the selection and balance of different methods used to explain this event? There are of course many ways to cut up the concept of “type of event,” but I will focus on two dimensions in particular: whether the event is rare or common; and whether the event is brought about by the decision of individual leaders of groups or through macro social processes that no specific individuals control (I will later bring in a third dimension, namely, whether the individuals that may be responsible for an event E are aware ahead of time that different types of endogeneity may be at play). The second question comes out of the first, namely: Is the researcher primarily interested in explaining specific historical events, or in explaining the general effects that a particular causal variable will have as it varies across space and time? This question relates to the familiar distinction between causes-of-effects research and effects-of-causes research.2 Yet when we combine this distinction with the answers to the first question on the type of event, we can better understand why there is no one best way to approach methodology or mixed-methods research in the social sciences.

As I have argued in EIAW and in the introductory essay for this symposium, when events are rare in IR and CP, they are typically rare for a reason. That is, it requires a number of conditions to come together simultaneously for the event E to occur. Wars between great powers, as shown in the book, and social revolutions in CP, are two clear examples of such complex conjunctural causality. Consequently, although an “individually necessary, jointly sufficient” (INJS) bundle of factors A, B, and C is needed before event E comes about, different bundles of factors may also lead to E (multiple causality or equifinality). Yet once we’ve established that complex causality is at work, we also need to know whether the event E is the product of individual decision-makers such as leaders, as in great power war, or a function of changes in larger social variables, as with unanticipated social revolutions such as the English, French and Russian revolutions.3 In situations where individual leaders choose to bring about E, especially when E has potentially dire consequences for their societies and their own families, we will be naturally interested to know why they did so, that is, the reasons that propelled them into such momentous decisions.4 When, on the other hand, the event E comes about because of macro-historical changes in society, we will be more interested in establishing exactly how the sequencing of these changes led to the crossing of a threshold that brought about the rare event E.5

In the study of rare events, because these events involve complex conjunctural causality, we will typically be more concerned with explaining why specific events came about when they did (a causes-of-effect approach) than in seeing to what extent the variable of our pet theory explains changes in E in general, all else being constant (the effects-of-causes approach). Yet if these rare events are brought about by individual leaders, as is the case in this book, the answers to “why?” and “when?” can best be ascertained by a study of the internal thoughts, attitudes, and motivations of the decision-makers themselves. In many situations, as with studies of voting or purchasing behavior in the United States or other countries, surveys and interviews can give researchers access to these internal processes of decision-making.6 Given the nature of great power politics, however, where there are no surveys of

---

1 Copeland 2015. To save space, I will often refer to the book as EIAW and its trade expectations theory as TET.

2 Mahoney and Goertz 2006; Goertz and Mahoney 2012.

3 Skocpol 1979; Goldstone 1991.

4 None of the three critiques engages my point that analysts doing research on complex causality need to determine what causal role A, B, and C are doing within the INJS bundle that leads to E. Documentary work is particularly valuable here, since it can reveal whether a factor is propelling the leader to choose E (the “reasons why” he or she acts) as opposed to simply constraining or facilitating the leader’s behavior (72-73).

5 Pierson and Skocpol 2002

6 For an example of this, see Welzel and Inglehart 2007.
leaders and interviews are either impossible or unreliable (74), one must focus on an intensive investigation of the documents that show what leaders were thinking as they changed their behavior (either toward war or risky hard-line behavior or toward greater cooperation). In the study of macro phenomena such as revolutions, documents are less important since the events are driven by the larger changes in society, not by the conscious choices of individual leaders.7

The above discussion suggests a 2 by 2 framework that lays out four main ways to approach the study of different types of events (what I will call for convenience the Approaches to Mixed Methods framework, or AMM). On the vertical dimension of AMM, we have either a causes-of-effects or an effect-of-causes approach (CE versus EC). On the horizontal dimension, we can distinguish between methods that seek to investigate “internal phenomena” (IP) such as thoughts and motivations versus those that are interested in observing “external phenomena” (EP) such as changing social and class structures, population densities, resource levels, and the like. The first approach, investigating causes-of-effects through the study of internal phenomena (CEIP), is the primary method employed in the book. CEIP is recommended when the type of event being explained is both rare and the product of the conscious decisions of individual leaders of states or key social groups. Since rare events are likely a function of complex causal conjunctures, researchers will be most interested in establishing what particular bundle of INJS factors were at work for specific events in history. What separates CEIP from a causes-of-effects logic that relies on observations of external phenomena (CEEP) is the way the events come about—by individual decisions of leaders rather than by a mix of social aggregates. In the book, crises/wars or the ending of cold wars are driven by choices of leaders, and thus we need to look “inside the heads” of leaders and their support staff to figure out why they made a dramatic shift in behavior. But for things such as unexpected social revolutions, our explanations rest on the general social conditions, not on the decisions of leaders per se. Here, the CEEP approach of such scholars as Skocpol8 and Goldstone9 thus makes sense.10

Both effects-of-causes approaches are, by definition, looking to explain the general effect of a particular variable over many cases. For a typical large-N study in IR and CP, one tends to use the fourth approach, namely, the observing of external phenomena such as trade levels, regime-type, and levels of development to establish the relative contribution of a particular causal variable A to the overall explanation of variation in event E. This is the ECEP approach to research. Yet, it is worth remembering that when reliable mass surveys of social phenomena are available, one can also do effects-of-causes research with internal phenomena (ECIP).11 Documents can also be useful in ECIP research, and indeed in my book I am also interested, as McKeown notes, in the general effect of changing trade expectations across many different kinds of scenarios. If I can show that rising or falling expectations of the commercial environment had a dramatic effect on the likelihood of conflict across many different scenarios, it suggests that TET’s main variable “A,” trade expectations, may be playing an important causal role within different INJS bundles that lead to war or to the ending of cold war, even if A is not implicated in all bundles that do so. Nothing in the discussion so

---

7 Indeed, in instances of this kind of phenomenon, the leaders want to avoid event E (the revolution). This fact means that documents will at most reveal the mistakes leaders made in not anticipating or countering the larger social changes that, in retrospect, researchers know led to revolution.

8 Skocpol 1979.
10 Note that for civil conflicts that are not spontaneous but rather the result of deliberate decisions by individuals leading distinct and well-organized social groups, the CEIP method may be useful, especially when combined with CEEP (thus the value of documentary analysis for the U.S. Civil War and the 1931-49 Chinese Civil War).
11 This is obviously a point most relevant to American Politics (AP) research but, as Inglehart’s work in CP shows, cross-country surveys can be quite revealing of the intervening social attitudes that can shape political outcomes (Inglehart 1990; Welzel and Inglehart 2007).
far indicates one must choose only one of the four approaches. The task is therefore to determine what relative balance one should maintain between the four different modes of investigation, depending on the type of event being studied and how one anticipates explaining it.

The McKeown and Gartzke Critiques

Let me now turn to the three critiques before us, so that we can see how they might contribute to a deeper understanding of mixed methods research in IR and CP—in particular, to knowing when and how to use emphasize certain methods, and with what relative “mix.” All three authors seem to accept that it might be worthwhile to investigate “how often” a variable such as trade expectations explains dramatic shifts within a case toward crisis and war or toward the ending of a cold war rivalry.12 In this sense, the effort to define a “universe” of cases, bounded by a time frame and the type of actor (great powers, for example), can be a valuable goal in qualitative work, since it allows us to see just how often one theory’s causal logic explains specific cases better than alternative arguments.13 Yet despite this agreed starting point, McKeown and Gartzke are concerned that I overemphasize the differences between qualitative and quantitative techniques, and the advantages of the former. Moreover, in accordance with Zaks, they also see my cases as being overly shaped by my consideration of what happens on the dependent variable. Zaks’s main concern, however, has to do with my initial set-up. Aside from the fact that she believes I undersell the true potential explanatory power of trade expectations theory, she suggests that I am uncertain as to what I am really doing with the theory. Sometimes I seem to be trying to use evidence to defeat rival theories (a “mutually exclusive” approach). At other times, she argues, I appear to be showing that different theories can both be “right” for specific cases, either because their variables work together (a “coincident” approach) or because the evidence could support competing arguments (a “congruent” result). In what follows, I will first deal with the McKeown and Gartzke critiques, and then turn my attention to Zaks’s more all-encompassing theoretical and methodological points.

McKeown argues that I use a combination of the congruence method and regular process tracing to establish the empirical value of my deductive argument. The congruence method focuses on the co-variation between changes on an independent variable and changes on a dependent variable within a case over time, or across cases. Process tracing differs from the congruence method insofar as it is more focused on identifying the intervening processes and variables that link a theory’s main independent variable or bundle of INJS conditions to the behavior or outcome of interest.14 McKeown is right that I employ both methods. But his analysis misses a crucial point: namely, that when the type of event being studied is a rare event that comes about through the conscious decisions of a few powerful men or women, the way one uses congruence or process-tracing methods is quite different from the way one studies non-rare events or macro-historical events. I am indeed interested in the overall co-variation between trade expectations and the probability of conflict between great powers and, in this sense, I want the theory to be relevant to those interested in the effects-of-causes across many cases. It is for this reason that I advocate analyzing “how often” a particular variable from a theory can offer some explanation for specific events over time and space. But because of the complex conjunctural nature of causality in great power conflicts, my main focus is on establishing explanations for particular phenomena, namely, dramatic shifts in behavior within designated case periods. Thus, the bulk of my work is devoted to CEIP analysis, the first quadrant of the AMM framework. In fact, we cannot establish “how often” a theory succeeds within an effects-of-cause approach until we’ve done the spade work on explaining the individual cases—which is determined by the type of event in question and the fact that it arises from complex decision-making of individual leaders and officials. In short, for rare events with individual decision-makers, CEIP must come before ECIP and is a necessary step for the latter.

Yet the charge that I have a distorted set of cases because of ill-advised selecting on the dependent variable—a claim all three authors make—misses what was said in the book about the unique nature of events that are both rare and driven by the conscious decisions of powerful elites. Complex conjunctural causality of this type means that elites, when rational, are fully aware of two fundamental forms of endogeneity: the risks of spirals and feedback effects when they act; and the ability they have to effect changes in other key conditions of state action. Gartzke thinks that I ignore the issue of endogeneity. The truth is the exact opposite: I deliberately incorporate endogeneity into the deductive logic of the theory, for the simple reason that individuals, who shape the major events of their nations, are only fully rational if they understand the likely effects of their actions on the future costs, benefits, and risks for their groups. Thus, the leaders of dependent states in my deductive set-up understand that if they are too aggressive in their efforts to control access to raw materials and markets, they can cause other states to cut them off, which might lead to a spiral of mistrust and hostility. They also know that if they can increase trade, this may make the other more inclined to peace, which in turn could lead to a further increase in trade (a virtuous cycle, rather than a negative spiral). Finally, they know that if they do need to go to war to protect their trade spheres, they must first “get their ducks in a row,” that is, they must adjust the conditions necessary for initiating war, e.g., building domestic and allied support and increasing arms spending.15

The realities of complex causality and endogeneity in the study of great power conflict means that in testing a theory that captures these realities, one necessarily must focus on

---

12 Copeland 2015, 77-78; 92-93.
13 Copeland 2015, 75-76.
14 See George and McKeown 1985; George and Bennett 2005.
15 Copeland 2015, 75 fn. 28. The importance of endogeneity and feedback loops is covered throughout the book (9-12, 39-43, 48-49, 74-75, 428-29), and is incorporated into a causal diagram summarizing the theory (49).
careful documentary process tracing of the major shifts in great power behavior, either toward crisis and war, or toward the ending of a long-term rivalry.\textsuperscript{16} What we need to know, above all, is whether the actors acted for the reasons hypothesized—did they take their nations into crisis/war or decide to end an dangerous rivalry because of changes in dependence and trade expectations, or for some other set of reasons? If our goal is to explain dramatic shifts in behavior, we can only know the reasons for the shifts if we look at internal phenomena, that is, at the beliefs, attitudes, and ends of the individuals making the decisions. In short, we must emphasize CEIP over the other the three methods, even if we supplement CEIP with them.

But why emphasize dramatic shifts in behavior as a basis for defining case periods? None of the three authors analyzes my stated criteria for defining a case period, or the reasoning behind them.\textsuperscript{17} Nor do they discuss the fact that I deliberately chose a “hard era”—the period from 1790 to 1991—for the testing of trade expectations theory.\textsuperscript{18} That being said, it is an

---

\textsuperscript{16} I should note for clarity that Zaks wrongly states that I only look at actual conflicts (crises or wars) as rare events E’s, and that I omit “non-conflict interactions.” Aside from the study of efforts to end the Cold War during the bipolar era of 1945 to 1991, I also discuss periods when states stayed peaceful in multipolar settings despite potential new reasons for crisis or war (Russia in the upheavals in the Eastern Mediterranean 1839-40 and Japan in the late 1920s, for example). As I state upfront, I am interested in changes in the probability of war due to dramatic shifts in behavior either toward crisis and war initiation, or due to unexpected efforts to end or calm a dangerous rivalry (vii, 3 fn. 2; see also 92, fn.37 on why most détentes in multipolarity are not examined).

\textsuperscript{17} In addition to using major shifts in behavior within an existing great power relationship or across regions as a basis for designating a case period, I also use the relative degree of independence of events as a criterion to ensure that years of struggle between great powers were not divided up too finely. As I noted, it would make little sense to treat the series of great power wars from 1803 to 1815 as “separate” case periods, since they were all shaped by one overarching fact: the hegemonic aspirations of Napoleon. Treating them as independent events would create a bias in favor of whichever theory did a good job for the 1803-15 period as a whole (giving it more “hits” than it deserved). A quick scan of some of my key case periods shows that my criteria for case periods almost always works against my theory, rather than for it. Since TET works well for periods such as the Napoleonic Wars, the 1900-14 period in Europe, and the 1930-41 period leading up to the Pacific War, counting these three periods as, say 10 or 15 separate cases would have greatly inflated the relative success rate of my theory versus competing theories. It is worth noting as well that I deliberately added in what might be seen as potentially marginal cases that do not work for TET such as the Belgium Crisis of 1830-31 and the French and Austrian interventions of 1815-22 in order to preempt charges of historical bias (91, 96).

\textsuperscript{18} As I pointed out, had I chosen the universe of cases for the 1550 to 1750 period, critics would have argued that TET and economic realism would have been more likely to work, since this was the heyday of mercantilism. The period after 1790, on the other hand, was increasingly informed by the liberal economic thought of Smith, Ricardo, and modern trade theory. The 1790-1991 era thus constitutes more than a fair test of realist-based theories (TET and economic realism) versus established liberal arguments (76). The fact that the latter does poorly even during the era of liberal economics is understandable concern that a book that focuses on explaining rare events such as crises/wars and ending of cold war rivalries might be downplaying, to use Zaks’s words, the regular periods of “stability” prior to dramatic changes in great power behavior. So aren’t there obvious distortion effects from emphasizing evidence based on the dependent variable? There are two ways to deal with such a question. The first is to admit that some distortion inevitably comes into play whenever research spends the bulk of its time explaining more “extreme” events,\textsuperscript{19} but then to argue that one has to try to minimize any distortions through certain techniques. In my case, I use two important techniques to avoid overstating the value of my theory or its causal significance. First, by examining the “universe” of cases for a bounded time frame (1790-1991) and type of actor (great powers), we automatically bring into the empirical analysis cases that have nothing to do with economic interdependence—hurting the overall explanatory power of any theory claiming a link between commerce and conflict. In this book, this was one-quarter or ten of the 40 case periods under consideration (92-93).

Second, for each case period, I deliberately spent time examining the periods of “stable” relations prior to the dramatic shifts in behavior that constitute the rare event. This allows us to see what shifts in independent variables led to a change in decision-making, and whether changing trade expectations were critical to the new decisions, and if so, in what way. Yet, as mentioned above, by using documentary process tracing and the CEIP approach rather than simply observing the arising of intervening processes through CEEP, we can understand why the actors changed behavior, as compared to the reasons they had for maintaining a stable relationship prior to that “decision point.”

In short, far from ignoring periods of “stability,” the study spends a great deal of time—especially for controversial cases where my evidence cuts against traditional accounts—in trying to understand the stability and what led to shifts in behavior within case periods.\textsuperscript{20} Notwithstanding this clarification, it is true that the book spends most of its time on the qualitative study of why leaders choose to initiate crises and war. Yet, all three authors miss the reasons laid out for such a focus. In any situation of complex conjunctural causality and rare events chosen by individual leaders, we would not expect a change from a “stable” pattern of behavior unless all the necessary conditions in an INJS bundle were met.\textsuperscript{21} Furthermore, the certainly a count against it.

---

20 For example, the case study of Japan’s shift from cooperation in 1920s to conflict and ultimately war in 1941 spans one and a half chapters (chs. 4-5), given the case’s complexities and the difficulty in showing the errors of traditional historiography. The book also looks at periods of tension that either did or did not shift to détente (1956-62, 1963-73, and 1984-91).
21 Note that stability is not defined here as “peace” per se, except insofar as peace means the absence of war. In most periods of no-war (not-E), great powers are in a low-level struggle for territorial and military position. They simply are not seeking costly wars or the crises that could lead to them.
absence of any one of these necessary conditions would lead to
the prediction that there would be no dramatic shift in behavior.22 Overall stability in a great power relationship, in short, is highly overdetermined. Moreover, since a leader who has decided to engage in war may take years to endogenously create the conditions necessary to execute it, what we really need to know is when the decision for war or crisis is made, and why. This requires an investigative focus on the “conflict” side of great power relations. For this to be done adequately, we must look at documents on the internal thinking and planning of the key officials of a country. For example, we cannot understand why World War II broke out in Europe in September 1939 without examining the period after January 1933—when Hitler and his military laid down and implemented a plan to get ready for major war within eight years.23

Gartzke, and to some degree McKeown, seem to eye suspiciously the effort to support trade expectations theory through the use of documents, as opposed to either large-N analysis24 or the process tracing of “publicly observable national features.”25 Gartzke suggests that this opens up the floodgates for the study of perceptions and thus misperceptions, and for subjective and cherry-picking analysis of what the documents say and mean. McKeown does not go quite this far, but he charges that much of my documentary process tracing relies on secondary sources and thus on the interpretations of historians who do not reveal their methods. This means, for McKeown, that my evidentiary base for conclusions is “indirect” and thus similar to my claim that large-N quantitative work can only “indirectly” get at expectations. The implication is clear: perhaps my documentary work is not so different from most qualitative process tracing and large-N quantitative work after all.

Seen within the full scope of what I was trying to do in the book, and my upfront efforts to specify how to deal with such potential problems, these critiques are off the mark. First of all, it is the very fact that expectations are a form of perception that forces us to see that we must study “internal phenomena” when the documents are available and reliable (CEIP over CEEP, and ECIP over ECEP as primary approaches). But I assume in their work I was considering.

One might argue that this is perhaps easy to do, since the notion that trade expectations drive behavior is almost completely ignored by historians, who tend to focus on domestic and ideological factors to explain big events. But that is exactly the point. Political science provides us with new theoretical lens to reexamine old events and to come up with new ways to explain them. Thus, it is up to Gartzke, McKeown, and other scholars to show, with evidence, exactly why my interpretations of evidence are indeed distorted, and why their interpretations are better. The fact that Gartzke and McKeown can neither point to specific examples of distortions in my case studies nor show how a different set of cases or different evidence would lead to contrary interpretations suggests that their charges of distortion are at this stage merely speculative.26 I ended my methodological chapter with a call for scholars to jump into the historical debates about specific cases and to correct my interpretations when errors are clear, thus help-

22 All three critics overlooked an important footnote in the book, drawing parallels between the study of rare events in the medical profession, such as the “Sudden Unexpected Deaths” (SUDs) of young people, and the method I advocate for considering rare events in IR and CP. SUDs are studied by taking a universe of actual deaths, to see “how often” the deaths were the result of asthma, heart failure, or stroke and the conditions that contributed to the event (such as hot days and excess exercise). While researchers may examine the days and weeks of the individual’s life leading up to the SUD, the focus on the INJS conditions for the SUD itself is critical, since a stable situation of “life” is so overdetermined for the average young person (see 78 fn.34).
23 Copeland 2015, 135-140.
24 See Gartzke’s essay for this symposium.
25 See McKeown’s essay for this symposium.
ing the field to build an increasingly accurate sense of how often different theories work and under what conditions. I suggested that only when this is done can we say that there is a growth of knowledge in the IR field.\textsuperscript{29} I still hold to this recommendation, and I encourage Gartzke, McKeown, and other scholars to participate in this long-term process of true knowledge accumulation.

My second point of rebuttal to McKeown’s implied charge that there is not much that is new here is more fundamental, since it concerns the larger question of how to do mixed-methods research well. Once we establish that we are dealing with a type of event that requires a deep investigation of the thinking and planning of leaders via a CEIP approach, we must delegate to a secondary status the CEEP process-tracing that focuses on changing “national features”. But as we’ve seen, there is nothing in my set-up that says one cannot employ all four approaches simultaneously as ways to reinforce the confidence in the findings that arise out of the primary method of CEIP. It is all a question of the relative mix or balance between them. For most of my cases, I employed documentary process tracing while also inserting observations of aggregate national features that would reveal the general social environment leaders were grappling with. For example, I brought in changing levels of democracy versus authoritarianism across many of the cases to show how these macro variables might constrain leaders or in fact facilitate or drive their plans for conflict or peace. This sounds like CEEP analysis, and it is. But it is very much secondary to CEIP precisely because of the type of phenomenon the book is trying to explain. Likewise, as mentioned above, I wanted to determine whether changing trade expectations shape conflicts across many different types of countries, regions, and historical periods. But my interest in establishing “how often” TET works well across the 40 case periods, while akin to a medium-N congruence test within the ECIP approach, is only done after the CEIP operations are performed for explaining specific cases.

Finally, I examined the “greatest hits” of the large-N quantitative work on trade and conflict, to see to what extent trade expectations logic can explain the results.\textsuperscript{30} This employs the fourth approach of ECEP in a fruitful and supportive way. By bringing in countries that are not great powers (and indeed make up most of the observations in such large-N studies), we can determine whether TET’s explanatory power is confined only to great powers. Moreover, using large-N can help to reduce any lingering anxiety that the results of the case studies are a function of the cherry-picking of evidence or of selection effects in general.\textsuperscript{31, 32}

We can see in this way that a mixed-method approach that examines many or all of the cells within the AMM framework can serve to reinforce confidence in findings that perhaps have come out of an emphasis on one of the cells. But for scholars of IR and CP, the question is always: Where should I devote my effort, given that time is limited and the skills required for doing each cell well are quite different and themselves take time to develop? The typical way to answer this question, at least in the IR field over the last two decades, has been to start with large-N quantitative tests and devote the bulk of one’s time there, if the data is reliable and abundant. King, Keohane, and Verba’s 1994 book (“KKV”), because of its bias toward causal inference based on quantitative methods, has for almost a quarter of a century reinforced this inclination of grad students and young scholars to see quantitative methods as superior, at least if good large-N evidence is available. Yet, as we have seen, when events are rare and driven by a few powerful leaders, documentary process tracing via CEIP must be the primary method to test competing causal logics (again, presuming reliable and ample documentation). Other methods, including large-N ECEP, may be useful supplements, but much is lost if they are made to be foundational.\textsuperscript{33}

The Zaks Critique

The Zaks essay offers quite a different critique of the book, which combines an effort to show that my theoretical argument is actually more all-encompassing than I recognize with an argument that the theory is inconsistently tested. Zaks offers useful advice to help correct the latter, namely, that a researcher should be clear whether his or her theory, relative to alternative arguments, is positioning itself as mutually exclusive, coincident, congruent, or inclusive. As I see it, Zaks’s essay seeks to show that once the imprecise elements of the current version of trade expectations theory, both deductively...

\textsuperscript{33} This is not the place to jump into the debate on causal inference. But it is worth noting that KKV and most studies challenging it simply assume that there is no real difference between documentary process tracing and process tracing involving observable social aggregates. Yet, in the latter, causal inference is indeed mere “inference” (informed conjecture), since one is only observing co-variation and inferring a causal relationship, say, by trying to hold as much as possible constant as one’s independent variable varies. As in the natural sciences, more cases are better—and experiments the apparent gold standard—since no single case or observation really says anything on its own (inference comes only from comparison). Yet in documentary process tracing, as in murder trials, one can use evidence on the internal thoughts and intentions of individuals in specific cases to, at the very least, eliminate possible causal explanations, even if one only rarely finds “smoking gun” evidence for a particular theory about event E. This gives us huge leverage when debating alternative causal accounts linking variables A, B, and C to a particular event E. The documents may show that C did co-vary with E, but that it played the role of constraining the individual to avoid E, rather than pushing them into choosing E (as pre-1914 domestic upheavals made German leaders wary about general war; see Copeland 2000: ch.3). Any theory arguing that C was the propelling reason for E could then be eliminated from contention. If one can use this approach to eliminate all but theory J, and J fits the evidence, the likelihood that J provides a good causal account of the arising of E is greatly enhanced.

\textsuperscript{29} Copeland 2015, 94–96.
\textsuperscript{30} Copeland 2015 53-69.
\textsuperscript{31} Copeland 2015, 3.
\textsuperscript{32} Gartzke disputes the way I interpret the large-N results of quantitative scholars studying trade and war (including his own results). Yet his criticisms amount to vague one-sentence dismissals rather than clear and sustained arguments.

\textsuperscript{33} This is not the place to jump into the debate on causal inference.
and empirically, are cleared up, the theory would have strong value as an extension of both liberal and realist thinking about the mechanisms linking economic interdependence to conflict or cooperation.

I greatly appreciate Zaks’s effort to make TET a stronger and perhaps “grander” theory of international relations. Overall, however, I believe Zaks pushes this agenda too far. If I were to adjust TET along the lines Zaks advocates, there would be much that would be lost, both for scholars and policy-makers. In particular, instead of having three or four big approaches (if we include neo-Marxism) explaining some of the cases some of the time, we would have only a single reformulated theory (TET) trying to explain almost all of the cases all of the time. As I will show, each of the competing arguments is indeed separate from TET, and by being so, brings something valuable to the theoretical and empirical table.

Using her Relationship Among Rivals framework, Zaks correctly points out that the book has some looseness of language that leads to confusion as to what trade expectations theory is seeking to do for the field of IR. She notes that at times I pose TET in a mutually exclusive way, as though it is a direct competitor to the alternative arguments of commercial liberalism and economic realism (and neo-Marxism). Yet, at other times, I suggest that variables from multiple theories might work together to explain particular cases (coincident explanations) or that they might be congruent in that the specific pieces of evidence might support multiple theories. Finally, Zaks indicates that I underestimate the value of TET, and that it in fact has “inclusive” power—namely, that it can and does incorporate the strengths of both liberalism and realism as it leaves behind their weaknesses. In recommending that I accept its inclusive strengths, she suggests that I break out of the political science habit of setting up mutually exclusive “competing” theories that seek to “beat” one another. Indeed, if TET is as powerful as she implies, it would probably be able to avoid empirical conclusions suggesting coincidence or congruence, since evidence that seemed to support liberalism or realism would really be evidence for what physicists would call “special cases” of the broader, more inclusive theory.

Let me start by saying that I quite like Zaks’s Relationships Among Rivals framework and believe it can help hone a researcher’s sense of exactly what he or she is doing. Indeed, I wish the framework had been available prior to the writing of the book—it would have helped me avoid falling subconsciously into language that implied that I was setting up a mutually exclusive empirical competition between TET and the alternative theories. Zaks’s observations on this language are accurate. But as the book’s methodology chapter and empirical cases make clear, I am above all what Zaks might call a “coincidence scholar.” That is, I believe strongly that—given the type of event I am studying, namely, complex rare events where individual leaders are conscious of the magnitude of their decisions both in terms of endogeneity and sheer social costs—different theories may indeed explain different aspects of a particular case, and thus all be simultaneously “right” at times. Across the 40 case periods, I regularly give two or more theories credit for capturing parts of the causal reasoning of leaders that pushed them into war/crisis or to end a cold war rivalry. Hence, for the 30 of 40 cases where commerce is involved in the shifts in behavior of great powers, the percentage of “successes” for the main three theories—TET, commercial liberalism, and economic realism—adds up to more than 100%. This of course would be impossible if establishing a mutually exclusive approach to testing had been my true intention.

Nevertheless, it is important to distinguish between a mutually exclusive approach to deductive theory building versus to empirical analysis. Because human beings are able to hold multiple motives in their heads simultaneously, it is more than likely in practice that they will sometimes act because of “mixed motives,” believing that a particular action will “kill two or more birds with one stone.” In the study of Nazism and the start of World War II, for example, I was more than willing to admit that Hitler and the German military leaders may have had more than one “reason” for initiating war in 1939. Who, after all, would want to deny that ideological objectives, national pride, and personal desires for glory—and not simply security concerns tied to Germany’s commercial dependence on other great power spheres—were important in such a case? When such “coincidence” occurs, the goal of researchers is to use documents and the sequencing of aggregate social changes to evaluate two things: what causal roles did the variables of different theories have and with what relative causal salience for the final result? Here, I believe, a careful use of documents and a study of the sequencing of social processes can reveal answers to these questions.

Yet coincidence in the real world does not require, or even suggest, that one should develop coincident deductive theories. In the book, I was careful to state exactly how TET differs in its causal logic from liberalism, economic realism, and neo-Marxism. TET follows economic realism by assuming that leaders in anarchic realms care only for the security of the state and society, and thus are worried about their long-term power positions in the system. In this sense, as I made clear, both TET and economic realism differ fundamentally from commercial liberalism, which assumes that leaders and social groups are driven by the overall utility/welfare they derive from various interests.

---

34 The four-fold RAR framework of Zaks, by this way of thinking, is really a four by two framework, since for each of her original four categories, one can distinguish between the deductive stage of theory-building and the more inductive stage of quantitative or qualitative analysis and testing.

35 Moreover, if we seek to stay objective, our answers can fruitfully adjust over time through debate and reflection. For example, in the case of World War II, my initial assessment that TET was the strongest theory (Copeland 1996) changed as I thought more about the documentary evidence and sequencing. At least for Hitler’s reasoning, I argue in the 2015 book, economic realism offers the superior explanation (133-42).

36 The four-fold RAR framework of Zaks, by this way of thinking, is really a four by two framework, since for each of her original four categories, one can distinguish between the deductive stage of theory-building and the more inductive stage of quantitative or qualitative analysis and testing.

37 Moreover, if we seek to stay objective, our answers can fruitfully adjust over time through debate and reflection. For example, in the case of World War II, my initial assessment that TET was the strongest theory (Copeland 1996) changed as I thought more about the documentary evidence and sequencing. At least for Hitler’s reasoning, I argue in the 2015 book, economic realism offers the superior explanation (133-42).
outcomes. This fundamental difference in assumptions about actor ends leads to very different causal reasons for why the trade environment might have positive or negative effects on the likelihood of great power conflict. TET and liberalism might both predict, in terms of co-variation, that higher trade might lead to a lowering of the likelihood of conflict. But while liberalism argues that this is due to trade’s ability to act as a constraint on preexisting domestic pathologies that might otherwise lead to conflict—the gains from trade increase the opportunity cost of going to war—TET would argue that a state growing in power through trade is likely to believe that continued peace enhances its long-term security. Liberals thus see a fall in trade as unleashing the preexisting domestic pathologies—meaning that the state is propelled into war by unit-level variables, not by security fears—while TET’s logic indicates that falling trade tied to negative expectations for future trade make leaders inclined to war for fear of a decline in power and a long-term loss in security.

By keeping the deductive logics of TET and liberalism quite distinct—“mutually exclusive” in theory, if not in practice—we can go to the documentary evidence and ask: Did the leaders make dramatic shifts in behavior for liberal reasons, or for the reasons expected by TET or some other theory? In some of the cases, such as World War II, we do see some evidence of “mixed motives” and thus coincidence. But for most of the cases analyzed, there was little or no evidence that falling trade unleashed preexisting domestic pathologies present in the initiating state. Rather, conflict was propelled by security fears for the future, either for economic realist or TET causal reasoning.

What about Zaks’s argument that TET is either congruent with economic realism or that TET incorporates it because, in Zaks view, economic realism does not have true variables, while TET does? Unfortunately, the first notion—that TET is congruent with economic realism, or with neo-Marxism, because all three theories might predict an increase in conflict when they observe an increase in trade levels—overlooks something crucial. As I explained in the book, even when theories make similar predictions that factor A may lead to event E, if the theories offer different reasons for why A leads to E, then the theories are distinct and can be tested as such. The problem of course for large-N work, or for process-tracing involving the observation of external phenomena (CEEP and ECEP), is that simply observing co-variation between A and E, or between A and intervening variable B and then E, will not tell us which theory best captures the causal reasoning of leaders for a particular case.

By unpacking the causal reasoning behind economic realism and neo-Marxism, we see that despite some similarities with TET’s causal logic, they are indeed separate and potentially powerful theories with real variables and should be treated as such. Economic realism starts with an assumption that in anarchy, great powers must assume the worst about adversaries. Given this, as a particular great power becomes more dependent on trade flows, its sense of vulnerability to cut-offs rises, and it will seize new opportunities to expand militarily to increase control over access, thereby reducing this vulnerability. Thus, increasing dependence levels and heightened opportunities for expansion are its core independent variables. Both economic realism and TET are “realist” in their assumptions that anarchy and uncertainty, plus the search for security, will lead great powers to worry about long-term power. But TET assumes that states do not assume “worst case,” and thus are sensitive to changes in the likelihood that other great powers will trade with them into the future (the “expectations of future trade” variable is thus key). TET also incorporates the defensive realist notion that endogeneity matters—that aggressive policies by one state may lead others to cut that state off from trade, which in turn may only increase the state’s willingness to expand militarily. So, while economic realism predicts more conflict when opportunities to reduce dependence by expansion arise, TET predicts that great powers will only expand when new threats to commercial access arise and when these threats cannot be reduced by diplomatic measures.

What we see from all this is that the way the different theories are specified will drive the way they are tested. Despite some admittedly loose language, the methodological agenda set forth in EIAW is one of laying out the distinct causal logics for the various deductive theories—to show under what conditions they would expect factor A to lead to event E or to not-E. Because there is often overlap in the predictions of the theories—TET and liberalism both predict falling trade will lead to conflict, while TET and economic realism/neo-Marxism predict increasing trade can also lead to conflict.

The same holds for liberalism versus TET. Zaks contends that TET can absorb the insights of liberalism once it properly captures such things as liberalism’s use of the shadow of the future. Yet, as I just reiterated, the two theories’ starting assumptions regarding actor ends are fundamentally different. Note also that liberalism’s shadow of the future variable is held constant in TET by the assumption of non-myopic rational actors with no “discounting” of the future (36 fn.38). TET, in short, holds unit-level variables affecting estimates of the expected value of future trade constant, while varying the likelihood of the other being willing to trade, while liberalism, consistent with its unit-level focus, allows for myopic actors with high discount factors, even as it stipulates that current cooperation will continue if non-myopic rational actors have clear means (including institutions) of punishing defections (17, 37).

44 Neo-Marxism predicts that great powers will be more expansive when trade rises, because key economic elites within a country fear a future cut-off from access to cheap raw materials, markets, and places to invest surplus capital (22-23). Neo-Marxism’s assumption that great power decisions for conflict are driven by pressure from economic elites out for profit, not national security, makes it a distinct theory whose causal reasoning can be tested against the documents (I found only a couple of cases where this causal mechanism had any real empirical support).
the only way to sort out which theory or theories does well for particular cases is to look at the documents. Observing co-variation between A and E/not-E through CEEP and ECEP will not suffice, because the causal reasons for action remain hidden.\(^45\)

In the end, TET, commercial liberalism, economic realism, and neo-Marxism remain distinct theories that resist efforts to create overarching hegemonic ("inclusive") approaches, and this is a good thing. Having distinct theories allows us to capture more fully the reality that in situations of complex causality, there are almost always multiple possible paths to rare events such as war. It also allows our documentary analyses to reveal the conditions under which different bundles of factors will indeed create a path to event E. Finally, it makes our theories and evidence relevant to real-world policy-makers. Such policy-makers need to know more than simply that the presence of a specific factor increases the overall risk of E by some percentage—the kind of thing effects-of-causes research, especially ECEP, is useful for. Rather, they need to know, for the specific case they are dealing with, which of the different theories best applies, given the conditions present. TET may do well across a majority of cases, but this does not mean one should apply it to every current real-world case. Theories that bring in the pathologies that may be present within dependent states, such as liberalism and neo-Marxism, may be more appropriate, even if they don’t have broad success across the larger universe of historical cases. The reasons why contemporary Iran or North Korea might launch a war should trade fall may be very different than why Japan attacked Pearl Harbor, and U.S. and European leaders need to be "armed" with different theories that facilitate a proactive comparison of the differences.

**Conclusion: The Way Forward**

This essay has sought to show that there is no one best way to test rival theories in IR or CP, nor is there one best way to balance or "mix" methods to maximize leverage and the causal understanding of political phenomena. It all depends on the type or nature of the event one is studying, and the type of causal relationship one seeks to investigate. Rare events of the kind that EIAW examines involve individual leaders of strong states that can not only choose to initiate crises and war (or to end cold wars), but also to spend months or years "preparing the ground" for a momentous state-level decision. In such situations, careful documentary process-tracing to uncover the thinking and planning of leaders and officials is the best means to establish what complex bundles of factors led to the decision to dramatically change behavior from the stable pattern of behavior that preceded it. In the language of the Approaches to Mixed Methods (AMM) framework laid out above, CEIP is the foundational approach and the others help to support or reinforce the findings. For other types of dependent variables, such as rare social revolutions or more common shifts in exchange rates and levels of development, simple comparative-historical process-tracing of macro social forces (CEEP) or large-N correlational analysis (ECEP) would be the primary techniques, with other AMM methods being used to build confidence in the purported causal mechanisms or to show how leaders can influence the broader trends in aggregate social forces.

The main questions that a researcher must ask him or herself prior to spending years learning methodological techniques and gathering and analyzing data are these: What am I studying, and what am I interested in finding out? The three critiques in this symposium help us hone our ability to answer such questions. They allow us to see that scholars must be clear in separating methods that look at co-variation in observable external phenomena, through process-tracing or quantitative analysis, from deep investigation of the thoughts, attitudes, and intentions of powerful individuals—and when both methods are used, to identify the strengths and weakness of each, relative to the type of event being examined. They also help us remember that the way we set up our theories deductively, and our clarity in defining assumptions, will likely not only shape whether we are engaging in a cause-of-effects or an effect-of-causes investigation, but whether we are building theories to "beat" the competition (mutual exclusivity) or are willing to accept that different theories might “work” for different aspects of a single case or event (coincidence). My book puts forward one way to establish a balance between different methods in a single research program. But the right "mix" for any mixed-methods research, even when the evidence for all four individual AMM methods is plentiful and reliable, will always, and inevitably, be unique to the study being undertaken.

**References**


Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures*. 

\(^{45}\) Process-tracing using the identification of intervening variables B and C between A and E/not-E can help, of course, but if done only through CEEP and ECEP, this method only exposes more levels of co-variation, begging the question of why A leads to B and C before leading to E/not-E. To stop this ad infinitum process, at some point documents and CEIP/ECIP will have to be used. See also fn. 33 on causal inference.
Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. Cambridge: Cambridge University Press.

2017 QMMMR Section Awards

Giovanni Sartori QMMMR Book Award

This award recognizes the best book, published in the calendar prior to the year in which the award is presented, which makes an original contribution to qualitative or multi-method methodology per se, synthesizes or integrates methodological ideas in a way that is itself a methodological contribution, or provides an exemplary application of qualitative methods to a substantive issue. The selection committee consisted of Hillel D. Soifer (Temple University), chair; Harris Mylonas (George Washington University); and Ron Krebs (University of Minnesota).


Prize citation, written jointly: The Politics of Resentment is a compelling account of the urban/rural political divide in contemporary Wisconsin. Among many excellent books under consideration, the committee found the book especially laudable for several reasons. One important innovation found in the book is the thoughtfully developed and precisely delineated concept of rural consciousness. The committee was also impressed by the depth and incisiveness of Cramer’s research, including multiple visits with the same subjects over a period of years. Her book provides a window into the current political moment, a deep understanding of a fundamental political divide and illuminates crucial aspects of American political identity. The book is especially noteworthy, however, for its lucid presentation of its ethnographic work, or what Cramer calls the “method of listening.” In clear language, Cramer explains her approach and discusses how she addressed methodological and practical dilemmas in the execution of her research. The result, as a model of research design and presentation, is a piece of scholarship that will become a touchstone of research methods classes.

Alexander George Article/Chapter Award

This award recognizes the journal article or book chapter, published in the calendar prior to the year in which the award is presented, which—on its own—makes the greatest methodological contribution to qualitative research and/or provides the most exemplary application of qualitative research methods. The selection committee consisted of Bear Braumoeller (Ohio State University), chair; Neta Crawford (Boston University); and Laia Balcells (Georgetown University).


Prize citation, written by the committee by Bear Braumoeller: This article explores a neglected topic, counterrevolution, in the context of the uprisings of 1848 and the subsequent restoration of order. In so doing, it provides insights relevant to the Arab Spring while highlighting the importance of cognitive heuristics and biases in revolution and selective learning in counterrevolution. The case study itself is outstanding both in detail and in theoretical relevance and illustrates the utility of well-crafted single-case studies. In scope, depth, and theoretical relevance it stands as an exemplary work of social science.
**Sage Paper Award**

This award recognizes the best paper on qualitative and multimethods research presented at the previous year’s meeting of the American Political Science Association. The selection committee consisted of Eva Bellin (Brandeis University), chair; Janet Lewis (U.S. Naval Academy); and Erica Simmons (University of Wisconsin-Madison).

**Winner of the 2017 Award**: Tasha Fairfield and Andrew Charman, “Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats.”

**Prize citation**: We are delighted to present Tasha Fairfield and Andrew Charman the Sage Award for best paper presented at APSA 2016. Their paper, “Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats” contributes to a new, growing area of work, applying Bayesian logic to qualitative analysis. Specifically, Fairfield and Charman argue that a Bayesian logical framework can contribute importantly to process tracing—especially when scholars disagree on inferences—and in doing so, can serve as an important bridge between qualitative and quantitative research. While Bayesian techniques are not appropriate for all qualitative work or all projects that employ process tracing, the paper does an excellent job building its case clearly and persuasively and shows how the technique can be useful for some research questions. The paper deftly attends to both technical and conceptual issues; it provides helpful, step-by-step guidelines on implementation; and it considers how use of Bayesian techniques may improve analytic transparency. Please join us in congratulating Fairfield and Charman on this important contribution.

**David Collier Mid-Career Achievement Award**

**Winners of the 2017 Award**: Tim Büthe, Hochschule für Politik/School of Governance at the Technical University of Munich and Duke University and Alan M. Jacobs, University of British Columbia.

**Prize citation, written for the committee by Peter Hall**: The David Collier Mid-Career Award was established in 2010 to honor a remarkable scholar who was also a founder of the qualitative and multi-method movement in political science. It is presented annually to recognize distinction in methodological publications, innovative application of qualitative and multi-method approaches in substantive research, and/or institutional contributions to this area of methodology. The award is made this year by a selection committee composed of (University of Texas at Austin), Elisabeth Jean Wood (Yale University), James Mahoney (Northwestern University), and Peter A. Hall (Harvard University, chair) and it goes jointly to Tim Büthe of the Technical University of Munich and Duke University and to Alan Jacobs of the University of British Columbia. They were nominated by a group of scholars, led by Kathleen Thelen and including Orfeo Fioretos, Walter Mattli, Lisa Wedeen, and Deborah Yashar.

Alan and Tim are so well-known to this Section that I need hardly list their achievements. For several years, they have been the editors of QMMR, the flagship publication of the discipline as Co-Chairs of the Qualitative Transparency Deliberations, a mammoth enterprise involving dozens of scholars designed to identify the issues that efforts to render research reports more transparent raise for scholars using multiple kinds of qualitative methods. Last but very far from least, Tim and Alan have both made important contributions to our understanding of qualitative and multi-methods in their own scholarship. These include Tim Büthe’s 2002 APSR article on “Taking Temporality Seriously” as well as the methodological sophistication of his prize-winning 2011 book with Walter Mattli, *New Global Rulers*, as well as Alan Jacobs’ important 2015 APSR article on Bayesian approaches to mixed methods and his lovely chapter on process tracing and ideas in the recent book on process tracing edited by Andrew Bennett and Jeff Checkel as well as the methodological sophistication of his prize-winning book, *Governing for the Long Term*. In short, there could be no more deserving recipients of this award.

Second, Tim and Alan went beyond the call of duty as editors of the newsletter to lead a major effort to consider the recent drive towards greater research transparency in the discipline. Along with the QTD steering committee, they launched and oversaw a nearly 18-month process to deliberate over the implications of this move for a wide array of approaches to qualitative empirical data collection and analysis in the discipline. Housed on a web platform, the QTD entailed a multi-stage process aimed at providing an inclusive outlet for scholarly discussion and debate. In the first stage, the QTD consisted of open discussions aimed at setting the agenda for more in-depth and focused deliberations. This initial step led to the formation of 13 different working groups focused on particular themes or methodological approaches that emerged out of the open discussion format. In the second stage, these working groups, each with its own online discussion board on the QTD platform, led virtual consultations and deliberations moderated by 42 scholars of diverse backgrounds from large and small colleges and universities both within and beyond the United States. Based on these exchanges, each working group drafted a report, which was posted on the QTD discussion board and open to comments until mid-November of last year. At present, the members of the various working groups are revising their reports to incorporate the feedback they received through the QTD portal and through a series of panels at the APSA 2017 meetings. The work of the QTD is a broad, collective effort and all contributors—whether committee members or commentators—should be applauded for their input. We owe a special debt of gratitude to Tim and Alan, who devoted over a year of their time to oversee this important pro-
The exemplary process they have spearheaded is a testament to the encompassing spirit of the QMMR section. As an open venue for deliberation about the potentially momentous shifts of the drive towards research transparency for some corners of political science, the QTD has provided a vital service for the discipline as a whole.

The QMMR section continues to attract a broad membership. With about 740 members, the QMMR is the second largest section of the more than 40 sections in the APSA. At the APSA meetings last August, QMMR sponsored five different short courses on both well-established and novel methodological approaches. (Full information on these courses is available on the APSA QMMR website.) After the elimination of graduate student membership fees as of January 2018, a change adopted in response to a proposal from the APSA Graduate Students Association, we hope to welcome even more young scholars working in the qualitative and multi-methods traditions.

The QMMR section is distinguished by the rigorous scholarship of its members, who collectively represent a wide range of methodological and theoretical approaches. The QMMR commitment to intellectual diversity and open dialogue and debate helps to foster an inclusive ethos, which can only strengthen the discipline overall.

Melani Cammett
mcammett@g.harvard.edu

Letter from the Editors

With this issue, we complete our term as co-editors of Qualitative and Multi-Method Research. It has been an honor to edit QMMR, which we had the good fortune of taking over as a well-established venue for cutting-edge work and high-quality debate about qualitative and mixed-method research in political science, thanks to the splendid stewardship of the editors who preceded us: John Gerring (2003-2006), Gary Goertz (2006-2011), and Robert Adcock (2011-2014).

We have sought to maintain not just the high standards established by our predecessors, but also the commitment to make and keep QMMR a publication that reflects the methodological pluralism of the QMMR community and appeals to a broad range of QMMR section members. The conversations that have unfolded across the articles and symposia published during our tenure leave us optimistic about both the possibility and the fruitfulness of civil and respectful intellectual engagement across sometimes profound ontological, epistemological, methodological, and subfield divides—even when the underlying disagreements themselves cannot be reconciled.

These conversations have been made possible by the many colleagues who have contributed to QMMR over the last three years, with whom we have greatly enjoyed interacting and from whom we have learned a great deal. Thank you!

We also thank the section and the Consortium for Qualitative Research Methods (CQRM) for their support; Allison Forbes (vols. 13 & 14), Andrew McCormack (vol.15:1) and Alberto Alcaraz Escárcega (vol.15:2) for their copyediting assistance; and Joshua Yesnowitz, QMMR’s longstanding production editor, for his many years of service. And we are grateful to Sebastian Karcher and Sarah-Anna Hogan of CQRM for doing the bulk of the work to bring QMMR more fully into the digital age: Each essay now has a unique “digital object identifier” (DOI), with links from each issue’s online table of contents to an open-access post of the essay. This change makes QMMR content more findable, accessible, and citable, and we hope that it gives QMMR authors and publications additional, much-deserved attention.

We have, during our time as editors, sought to open QMMR to a broader range of authors by scanning conference programs for innovations and debates, of which we would otherwise have been unaware, and by inviting, reviewing and publishing unsolicited submissions. That said, looking back on volumes 13 through 15, we realize that our efforts to diversify QMMR authorship clearly paid insufficient attention to gender balance. The underrepresentation of women, who have comprised only 15% of the authors in QMMR during the last three years, is a serious problem that we regret we did not do a better job addressing.

All the more, we are absolutely delighted that Jennifer Cyr and Kendra Koivu have agreed to take on the QMMR editorship, starting with the next issue. Jennifer and Kendra have already been shaping qualitative and multi-method research in our discipline for many years, through (among other things) their publications and their roles as co-founders and co-organizers of the annual Southwest Workshop on Mixed Methods Research. We could not imagine this publication being in better hands, and look forward to seeing the directions in which Kendra and Jennifer will be taking the QMMR conversation.

Tim Büthe
buthe@duke.edu

Melani Cammett
mcammett@g.harvard.edu

Alan M. Jacobs
alan.jacobs@ubc.ca
Qualitative and Multi-Method Research
Department of Political Science
140 Science Drive (Gross Hall), 2nd Floor
Duke University, Box 90204
Durham, NC 27708
United States of America

Qualitative and Multi-Method Research (ISSN 2153-6767) is published twice yearly on behalf of the American Political Science Association’s Organized Section for Qualitative and Multi-Method Research. It is edited by Tim Büthe (tel: 919-660-4365, fax: 919-660-4330, email: buthe@duke.edu or buthe@hfp.tum.de) and Alan M. Jacobs (tel: 604-822-6830, fax: 604-822-5540, email: alan.jacobs@ubc.ca). The production editor is Joshua C. Yesnowitz (email: jcyesnow@bu.edu). The manuscript editor is Andrew McCormack. Published with financial assistance from the Consortium for Qualitative Research Methods (CQRM): http://www.maxwell.syr.edu/mynihan/cqrm/About_CQRM/. Opinions do not represent the official position of CQRM. After a one-year lag, past issues will be available to the general public online, free of charge, at http://www.maxwell.syr.edu/mynihan/cqrm/Qualitative_Methods_Newsletters/Qualitative_Methods_Newsletters/. Annual section dues are $9.00. You may join the section online (http://www.apsanet.org) or by phone (202-483-2512). Changes of address take place automatically when members change their addresses with APSA. Please do not send change-of-address information to the newsletter.