Document Delivery Service

ANALYTICAL ESSAY

Tipping Points: Challenges in Analyzing International Crisis Escalation

Chong Chen [iD]
Tsinghua University, China

Jordan Roberts
Coastal Carolina University, USA

Shikshya Adhikari and Victor Asal
University at Albany, USA

Kyle Beardsley [iD]
Duke University, USA

Edward Gonzalez
University of Southern California, USA

Nakissa Jahanbani [iD]
United States Military Academy, USA

Patrick James
University of Southern California, USA

Steven E. Lobell
University of Utah, USA

Norrin M. Ripsman
Lehigh University, USA

Scott Silverstone
United States Military Academy, USA

AND

Anne van Wijk
University of Southern California, USA

Why do some near crises tip over into full-blown crisis and others do not? This paper considers existing scholarship and identifies four key barriers to using quantitative analysis for tipping-point analyses: strategic indeterminacy; the incentives for conflict parties to avoid inefficiencies; the paucity of cases; and the availability of quality data. Due to these challenges, many do not perform well as immediate causes for crisis escalation. We also argue and demonstrate through two quantitative models of crisis

escalation that some variables, particularly related to domestic politics, can do well in explaining why some disputes tip into crisis and others do not. As we illustrate with reference to the 1995–1996 Third Taiwan Straits Crisis, qualitative approaches that analyze the processes by which leaders and foreign policy institutions make decisions add needed explanatory power to purely quantitative models of the potential for near crises to tip into crisis.

¿Por qué algunas crisis menores dan lugar a una crisis a gran escala y otras no? En este artículo, se analizan los estudios existentes y se identifican cuatro obstáculos clave al utilizar el análisis cuantitativo en los análisis de los puntos de inflexión: la indeterminación estratégica, los incentivos de las partes en conflicto para evitar las ineficiencias, la escasez de casos y la disponibilidad de datos de calidad. Debido a estos desafíos, muchos no se desempeñan bien como causas inmediatas de la escalada de crisis. Además, argumentamos y demostramos, mediante dos modelos cuantitativos de escalada de crisis, que algunas variables, en particular las que se relacionan con la política interna, pueden brindar una mejor explicación sobre por qué algunas disputas dan lugar a una crisis y otras no. Al referirnos a la tercera crisis del estrecho de Taiwán entre 1995 y 1996 como ejemplo, los enfoques cualitativos que analizan los procesos mediante los cuales los líderes y las instituciones de política exterior toman decisiones aportan el poder explicativo necesario a los modelos puramente cuantitativos de la posibilidad de que las crisis menores den lugar a una crisis a gran escala.

Pourquoi certaines quasi-crises basculent-elles dans une crise totale alors que d'autres non ? Cet article examine les recherches existantes et identifie quatre obstacles clés à l'utilisation de l'analyse quantitative pour les analyses de points de basculement : l'indétermination stratégique, les motivations des parties en conflit à éviter les inefficacités, l'insuffisance des cas et la disponibilité de données qualitatives. Du fait de ces difficultés, beaucoup de ces analyses ne sont pas très efficaces lorsqu'il s'agit d'identifier les causes immédiates de l'escalade d'une crise. Nous soutenons et démontrons également par le biais de deux modèles quantitatifs de l'escalade des crises que certaines variables, en particulier celles qui sont associées à la politique intérieure, peuvent s'avérer efficaces pour expliquer pourquoi certains conflits basculent dans une crise alors que d'autres non. Comme nous l'illustrons par des références à la troisième crise du détroit de Taïwan de 1995–1996, les approches qualitatives qui visent à analyser les processus par lesquels les dirigeants et les institutions de politique étrangère prennent des décisions ajoutent le pouvoir explicatif nécessaire aux modèles purement quantitatifs du potentiel de basculement des quasi-crises dans des crises.

Introduction

Recent years have seen a growth in the availability of granular data related to the behavior of states and other armed actors, as well as in the development of more sophisticated methods to analyze those data (e.g., Muchlinski et al. 2016;

Colaresi and Mahmood 2017; Hegre et al. 2019; Williford and Atkinson 2020). Despite these advances, studies intended to accurately predict the escalation of international crises face fundamental challenges. Even the best models produce high numbers of false positives or false negatives (or both). Poor predictive accuracy, in turn, raises questions about whether the component variables have much explanatory power in anticipating conflict escalation and de-escalation (Ward, Greenhill, and Bakke 2010; Schneider, Gleditsch, and Carey 2011).

This paper considers existing scholarship and highlights why many variables commonly used in the study of conflict actually do poorly in anticipating whether an international crisis escalates.[1] Strategic indeterminacy, the incentives for conflict parties to avoid inefficiencies, the paucity of cases, and the availability of quality data all limit the ability for models to accurately explain the cases that tip into crisis (and possibly beyond) and the cases that do not.[2] Our approach builds on the existing distinctions between permissive conditions of conflict and immediate causes of conflict.[3] In doing so, we argue and demonstrate that many variables used in quantitative models face substantial limitations in being able to explain the immediate causes or triggers of crisis.[4]

Using the canonical Cuban Missile Crisis as an example, variables typically found in quantitative models of conflict behavior such as those related to regime differences, alliance patterns, and power balance can help explain why relations were tense among the United States, USSR, and Cuba in 1962, but they do not explain why Khrushchev decided to deploy the missiles or why the United States responded with an ultimatum and "quarantine" of Cuba. They serve well as permissive but not immediate causes of crisis. To have anticipated the escalation of a crisis in October 1962, we would have needed to consider variables related to the political pressures on Khrushchev, Kennedy, and Castro; the reputational effects of previous crises such as the Bay of Pigs fiasco and the 1961 Berlin crisis; the extent to which the missiles affected the existential threats facing the regimes in Washington, Moscow, and Havana; and tendencies related to the psychology and bureaucratic decision-making of the key principals among others.

As another example, drawing from the Third Taiwan Straits Crisis (1995–1996) discussed below, we can understand the permissive conditions for conflict in that case, including rising Chinese economic and military strength, the security ties between Taiwan and the United States, and the history of tensions among these three actors. None of these factors, however, explain the outbreak of a *crisis* in July 1995. As we discuss below, the immediate causes center on the politics within the Chinese Communist Party (CCP) in early 1995, which incentivized President Jiang Zemin to take a hardline stance against tightening US–Taiwan relations.

Due to the four challenges—strategic indeterminacy, the incentives for conflict parties to avoid inefficiencies, the paucity of cases, and the availability of quality data—many structural variables, including those related to rising and declining states and their coercive power, will tend to do poorly in explaining why—and thus anticipating whether—near crises tip into crisis.[5] Even event-level variables related

---

[1] Our primary interest is in the crisis-escalation explanatory power of particular variables, and, following Ward, Greenhill, and Bakke (2010), poor predictive performance is an indication of poor explanatory performance.

[2] An international crisis represents a disruption in international processes and a challenge to the structure of the international system or one of its subsystems (Brecher and Wilkenfeld 2000). We specifically focus on models of the occurrence of an international crisis between two or more states, among which there is perception of a threat to basic values, a finite time for actors to respond, and a heightened probability of military hostilities (Brecher and Wilkenfeld 2000).

[3] Indeed, systemist approaches have begun uncovering ways in which scholars can bring together permissive and immediate causes into the same theoretical framework. See Bunge (1996) and James (2019).

[4] Permissive causes are the variables in which variation can help anticipate a heightened probability, or risk, of a crisis, while immediate causes are the variables in which variation can help anticipate the actual triggering of a crisis among the cases that are at high risk.

[5] See Iakhnis and James (2019) for an overview of the near crisis concept as a period of tension short of crisis.

to the actions and postures that states take in their efforts to signal the credibility of their threats are likely to do much better as permissive indicators of crisis rather than as immediate indicators of an imminent crisis. Conditional on a case having arrived at a tipping point, quantitative models face stark limitations in explaining why some cases tip into a crisis while others do not.

In light of these challenges, we make two additional points that stand to improve the explanatory power of tipping-point analyses and policy-relevant strategic assessments of escalation going forward.

First, variables related to leader incentives to avoid severe punishments for pursuing accommodations after initial aggressive posturing (Debs and Goemans 2010) have more potential to explain why some high-risk cases escalate while others do not. Gains in the construct validity and precision of variables used to measure the incentives and constraints faced by foreign policy makers could stand to substantially increase the explanatory power of our quantitative models.

Second, qualitative approaches face their own challenges in producing accurate explanations of crisis escalation, but they also provide advantages over solely quantitative approaches. Recent trends in data science portend that some of the limitations in models of crisis escalation will be reduced with further advances in data collection and statistical methodology, but some are unresolvable. Qualitative methods are in some contexts better suited for identifying why a particular near crisis tipped or did not tip into crisis. Policy-relevant accounts of crisis escalation, including national threat assessments in the intelligence community (IC), call for a combination of quantitative and qualitative methods.

Motivation

Our project is not the first to highlight the challenges of explaining conflict escalation and de-escalation. Even the most rigorous studies focusing on predictions of intrastate conflict have pointed out that models that rely on structural variables—factors that shape the environment in which the actors interact—tend to do poorly in predictive accuracy, especially when it comes to out-of-sample forecasting (Ward, Greenhill, and Bakke 2010; Gleditsch and Ward 2013; Chenoweth and Ulfelder 2017; Chiba and Gleditsch 2017; Blair and Sambanis 2020; Bowlsby et al. 2020).

Almost by definition, structural variables help anticipate permissive conditions for conflict, yet cannot predict well if the actors will escalate their conflicts when the permissive conditions are present.[6] Models that rely on structural conditions will do better in anticipating the absence of conflict; few cases will escalate when such models did not identify them as high-risk cases. Indeed, some cases may never experience the outcome of interest because various necessary conditions are not present, which can be modeled with knowledge of various structural conditions (Beger, Dorff, and Ward 2016).[7] However, models with only structural variables struggle in distinguishing between false positives and true positives—once actors experience an elevated risk of conflict, their strategic choices play important roles in determining whether a crisis will escalate or not. The relationships between structural variables and conflict escalation can also change over time, which makes it difficult for models to have enduring explanatory power (Bowlsby et al. 2020).

Studies that bring in data more related to the actor-specific incentives for escalation outperform models that lack such variables. Yet, even these research designs still produce many false-positive predictions. In an improved model of dispute initiation that Gleditsch and Ward (2013) offer, only 44 of the 125 high-predicted-probability (35 percent) cases actually experienced a dispute. That is, 65 percent of

---

[6] Waltz (1979, 121–22) makes a similar point in clarifying that his structural theory of international politics is not a theory of foreign policy. See also Schneider, Gleditsch, and Carey (2011).

[7] Models that focus on politically "relevant dyads" account for some of the factors that make it nearly impossible for conflict to occur between some states (Lemke and Reed 2001).

the cases that were identified as having a high risk for escalation did not actually escalate. Similarly, Blair and Sambanis (2020) develop a model of war occurrence with event-data information about escalation dynamics, which outperforms other models, and 132 of the 134 cases that were not at an elevated risk of experiencing war actually did not experience war.[8] While the false negatives are low, still seventeen of the thirty cases (57 percent) with an elevated risk of war did not experience war. Bowlsby et al. (2020) used a model of instability from Goldstone et al. (2010) and found that the model predicted only six out of seventeen instances of instability in the 2005–2014 period. Hegre et al. (2021) evaluated the model of Hegre et al. (2013) over a nine-year out-of-sample period and found that while the overall true positive rates of armed conflict onset are greater than 50 percent, the model performs much worse in anticipating low-level armed conflict onset.

If our best models still produce estimates of escalation in which fewer than 50 percent of the anticipated cases actually escalate when they are expected to be at an elevated risk, they remain limited in their explanatory power and in their utility for policy makers. We turn to four aforementioned challenges that existing scholarship points to as impediments to the explanatory power of our quantitative models.

Challenges to Explaining Tipping Points

Strategic Indeterminacy

The first challenge to using many variables to explain why an at-risk situation tips into a crisis or not stems from an understanding of the actors involved as strategic actors. Actors are not likely to take offensive or defensive actions that can be easily anticipated for fear that the adversary will deploy an effective counterstrategy that will obviate any potential benefit of taking the anticipated offensive or defensive action in the first place. Surprise is a crucial element of military strategy (e.g., Luttwak 2001; Betts 2010), and what was surprising to the actors in conflict is likely to appear surprising to analyses that use past behavior to anticipate behavior in a given moment. One implication is that analysts may find it difficult to connect the observation of a particular action or observed posture with expected outcomes. Actors will often take actions that are not well anticipated, so the choice of one action among the alternatives will have a strong stochastic component.[9] The observation of a particular action is not likely to be informative of whether the situation can be managed short of escalation or is heading toward a potential for escalated violence.

The problem can be reduced to a type of endogeneity problem in which the explanatory variables reflect strategic choices. Actors take actions with an eye toward the expected outcomes of the actions, which means that they will not take actions that they expect their opponent to effectively counter. Regression models of crisis escalation on action-level variables—for example, as measured by event data of (de-)escalatory statements and activities—will produce biased results because we should expect the data-generating process to be one in which the independent variables (reflecting actor choices) are correlated with the error term (reflecting expected outcomes). We do not get an accurate glimpse of the effect of the action because we do not observe the cases in which the actor chooses a different option in the same situation. Fearon (1994, 2002) applies a similar logic to explain the challenges of testing rational deterrence theory. A resolved defender of the status quo might take an action meant to deter a challenger, but we would then only observe challenges when the challenger has exceptional resolve of its own and thus when deterrence is most difficult. Since we do not observe all the cases of less-resolved challengers actually challenging when the defender takes a particular action, we do

---

[8] See also the subsequent debate between Beger, Morgan, and Ward (2021) and Blair and Sambanis (2021).

[9] Strategic indeterminacy is likely to increase with the stakes in play, as the incentives to counter an opponent's every move increase.

not observe the unbiased effect of the action on deterrence. Drezner (2003) similarly explains why we might not observe the effect of sanctions even if they can serve as tools of economic coercion. Signorino (1999), Smith (1999), Lewis and Schultz (2003), and Carrubba, Yuen, and Zorn (2007) describe how the nature of the estimator bias in crisis escalation models depends on the underlying form of strategic interaction. Models of crisis escalation not built on a correctly specified model of strategic interaction are thus likely to misestimate the actual effect that actor choices have on the potential for crisis escalation.

A separate omitted variable problem is also apparent when we consider the potential role of third-party conflict managers. If third parties are more likely to intervene when they interpret the protagonists' actions to reflect an at-risk situation likely to tip into a crisis, then the actions that otherwise would correlate with escalation will be observed to have a muted (or negative) effect because these are the cases that are more likely to have experienced a form of conflict management intervention (Gartner and Bercovitch 2006). The actors' strategic interaction dynamics thus make it harder to forecast when a near crisis is likely to escalate.

Incentives to Avoid Inefficiencies

Building on Gartzke's (1999) notion that "war is in the error term," if crises are costly to the protagonists—like wars though to a far lesser extent—then they have a vested interest in resolving their disputes peacefully.[10] Failure to do so, as a result of miscalculation or misestimation, should be stochastic and not well correlated with observable variables. If an actor knew that a particular observable variable was correlated with an opponent's reservation point—for example, if the variable was associated with an actor that was more resolved and intransigent—then the actor should anticipate it and account for that expectation in its demands and willingness to accept demands. Since demands are endogenous to expectations about what the adversary is willing to accept, actors should be more accommodating to adversaries who are more resolved, which means that the variables that relate to the willingness of an adversary to fight should affect the distribution of demands but not the probability of bargaining failure (Fearon 1994; Powell 1999).

Observable structural variables, such as those related to changes in power asymmetry, should thus have little explanatory power in anticipating the direction in which a tipping point will actually tip. Actors should take those variables into account when making their offers and counteroffers. They can help us understand the terms of settlement, and even the overall incentive of actors to try to renegotiate the status quo arrangement (Carroll and Kenkel 2019), but they should not much help us understand why bargaining fails and leads to crisis escalation in specific cases of renegotiation and not others.[11]

This challenge parallels those entailed in predicting and explaining financial markets. The market efficiency hypothesis posits that when financial markets are efficient, they should behave as a random walk process (Samuelson 1965). Just as investors have an incentive to buy and sell at a market price that accounts for all the information about the relative profitability of a company, so armed actors have an incentive to offer terms that account for all the information about what the adversary is willing to accept. Of course, financial markets do not always operate

---

[10] Indeed, some crises may not involve significant costs, at least in a material sense. The argument here applies more to crises that incur costs on the belligerents and less to crises that are not significantly more costly than concessionary steps that avert crisis. The argument here about avoiding inefficiencies is related to the problem of strategic indeterminacy in the sense that both can be couched in terms of endogenous choices. We present them separately in order to focus on action-level variables as germane to the problem of strategic indeterminacy and structural variables as germane to the problem of avoiding inefficiencies.

[11] As a partial exception, see Slantchev (2011) for a model of how stronger actors can undertake costly mobilization to prevail short of war in international conflicts.

efficiently, and so too do adversaries not always bargain efficiently; the sources of inefficiency are often instructive in pointing to causes of conflict (Fearon 1995). Periods in which the actors are at risk of operating inefficiently—as when there is a lack of information, a potential for misperception, irrationality, or constraints on policy making that reflects the interests of the country—are periods at heightened risk of conflict escalation.[12]

An understanding of the incentives for actors to avoid inefficient conflict bargaining further points to variables that potentially add to our understanding of tipping-point tendencies. In some cases, actors actually have an incentive to bargain inefficiently, as when escalation of a crisis would produce net benefits irrespective of the expected outcomes. The study of audience costs has shown that in some cases, leaders can get positive political gains for taking a hard line vis-à-vis an adversary, and there also might be heavy costs for making unpopular concessions (Debs and Goemans 2010; Croco 2011; Kertzer and Brutger 2016). Domestic factors that might affect a leader's willingness to take a hard line or to compromise include the expected severity of punishment (Debs and Goemans 2010; Goemans 2010); the degree of popular hostility to compromise; the stability of the regime (Ripsman 2016, 29–31); the government's degree of "stateness," comprising its autonomy capacity and legitimacy (Blanchard and Ripsman 2013); or the leader's tolerance for risk (Holsti 1972; Horowitz, Stam, and Cali 2015). To the extent that domestic variables can account for when leaders find crisis escalation as a valuable end in itself, these variables can help us anticipate the situations in which leaders will take their states headlong into crisis situations.

Too Few Cases

A third problem is that, unlike other domains in which machine learning algorithms have greatly improved the predictive accuracy of quantitative models, there are too few instances of international crisis to make much use of data-hungry inductive algorithms. The International Crisis Behavior (ICB) project (Brecher and Wilkenfeld 2000; Brecher et al. 2020) currently has coded information on 487 international crises from 1918 to 2017,[13] but the possible combinations of variables that could be used to explain crisis onset are virtually infinite. We quickly run out of degrees of freedom in trying to explain crisis onset, such that few cases of potential crisis onset closely resemble any of the cases that have been examined previously. Statistical forecasting approaches rely on the existence of analogous situations, so that the characteristics of the current period can be compared to similar cases in the data. If many factors matter to a predicted behavior such as crisis onset, it will be common for there to be no analogous cases of comparison to the current period and thus larger errors in prediction. The problem of finding analogous cases is compounded when considering that the most important causal relationships that shape conflict and peace vary from period to period. Jenke and Gelpi (2017), for example, find that the variables that are most important for explaining the onset of war in the post-World War II era are different from the variables that are most important for understanding onset in the interwar period, which, in turn, are different from the most important variables to explain the pre-World War I period. In this way, a present-day crisis might be hard to predict if the most recent analogous near crisis between adversaries occurred long ago.

This problem of too few cases relates to the problem above about strategic indeterminacy. Crisis escalation rises from complex and strategic interaction, and the

___
[12] It is important to recognize, however, that these conditions are permissive conditions and still leave us with the challenge of modeling escalation when the actors are already at the tipping point.

[13] More information about the current version (Version 14) of the ICB project can be found at https://sites.duke.edu/icbdata/. Accessed July 4, 2021.

existing level of theoretical development is not such that we have sufficient ex ante expectations about how to parameterize a model that well approximates the data generating process of such complex interactions.[14] At the same time, we do not have enough data to expect purely inductive forms of inference to identify variable combinations that well explain the expected outcomes under each condition in out-of-sample data.[15]

<div style="text-align:center">

Data Availability and Quality

</div>

A final challenge to developing adequate models of tipping points relates to the problem of available quality data. Armed actors have an incentive to keep information about their intentions, capabilities, and resolve secret, which means that much of the information related to whether the actors will choose to escalate or not is unavailable to the researcher. This problem is compounded for more recent cases or for cases involving closed regimes, in which archival materials are not readily available to the researchers (Ripsman 2022). The greater the distance between the information that the decision maker has and the information that the empirical model has, the more challenging the ability to form accurate expectations.

In part because of challenges in being able to observe the variables that should best capture escalation dynamics, quantitative analysis of crisis escalation faces prodigious construct validity problems. The variables that are coded regularly and consistently often only resemble crude proxies of the underlying concepts (Blair and Sambanis 2020; Ripsman 2022). Even if our theoretical understanding of crisis escalation improves, it may not be possible to develop better empirical models because the available data are not well mapped onto the concepts in the theoretical models.

Furthermore, data missingness problems are rampant. Even very basic data are often not available for all cases of interest. Consider, for example, the oft-used operationalization of economic interdependence: trade as a percentage of GDP (e.g., Oneal and Russett 1997). While available from the OECD, that dataset does not include data on trade between non-OECD members (Ripsman 2022). Researchers are often left with the choice of using incomplete data, which could create sample-selection bias, or restructured data, which could create measurement bias, or only including variables that do cover the entire set of relevant cases, which then could create omitted variable bias.[16]

In addition to fundamental challenges of measuring the concepts that matter, other challenges to data quality exist in the use of event data. These problems include inconsistencies in news coverage areas and event duplications. These challenges are well known (Beger, Morgan, and Ward 2016; Blair and Sambanis 2020) and less interesting to the study at hand because the problems and thus solutions are technological.

A related problem is the practical problem of gathering data. Accurate data collection from secondary sources often requires contextual knowledge (i.e., which journalists are knowledgeable about a specific country's policy preferences, which outlets are government mouthpieces, etc.), and accurate data collection from primary sources often requires access to and understanding of non-English language text. Few project managers and their research teams have sufficient capabilities or

---

[14] See Bueno de Mesquita (2011) for a type of model that has shown relative promise by combining formal theoretic models with expert coding of some of the key variables.

[15] In this vein, Cederman and Weidmann (2017, 475) note, "In the absence of full knowledge of how all theoretical components interact and sufficient data to measure the relevant variables, all that can be hoped for is risk assessment on the basis of structural features that increase the probability of conflict."

[16] Advances in imputation methods (e.g., Hollenbach et al. 2021) have helped reduce these trade-offs, but imputation trades inefficiency for bias, especially in the absence of alternative variables that are well correlated with the ones of interest.

resources to systematically collect large volumes of data with high accuracy and precision across many cases, especially when the measurements are intersubjective (Gleditsch, Beardsley, and Polo 2010).

Challenges related to data availability and quality speak to an important difference between statistical models that are used for explanation and those that are used for prediction. As-if-random measurement error is not necessarily a problem for explanation and causal inference, but it is a nontrivial problem for prediction (Shmueli 2010). Large measurement error will not systematically bias the direction of the observed marginal effect between an independent variable and the dependent variable, although it can produce attenuation bias and thus more conservative hypothesis testing. So, hypothesis testing is still possible when there is substantial measurement error, as long as the measurement error does not correlate with the explanatory variables or dependent variable. However, large measurement error necessarily makes forecasting more difficult, as noisy inputs lead to noisy expectations.

Modeling ICB Crisis Triggers: Starting with a Quantitative Approach

To illustrate the challenges facing models of tipping-point behavior, we use data from the ICB project (Brecher and Wilkenfeld 2000; Brecher et al. 2020). We employ two models—a machine learning algorithm (random forest model) and a parametric model (split-population duration model [SPDM])—to assess which variables explain the monthly variation in crisis escalation. We expect that structural variables related to changes in relative power as well as action-level variables will do poorly in explaining the occurrence of crisis among the set of cases with an elevated risk. Drawing on the work of Debs and Goemans (2010), we further expect that variables related to domestic incentives to stay on a belligerent course should have more potential to explain crisis onset and should condition the explanatory power of action-level variables.

We recognize the distinction between empirical approaches that try to proffer causal explanation and approaches that try to predict (Shmueli 2010; Cederman and Weidmann 2017). Since explanations that do not help us better anticipate tipping-point behavior are not helpful for the policy-relevant questions at hand (Ward, Greenhill, and Bakke 2010), we assess explanatory power as a function of both statistically significant associations and predictive accuracy.[17]

We focus on all *nondirected dyad-months* in the international system from 1995 to 2017, excluding those dyads that did not have any events in any months in the Phoenix event data (Althaus et al. 2020).[18] The dependent variable in the models is a dichotomous variable for whether an ICB crisis occurred in a given month between the actors in each dyad.[19]

Explanatory Variable Measurement

We include a number of explanatory variables. The first set of variables measures aspects of the states' *rise and decline* in power. The degree to which states can

---

[17] See Brandt, Freeman, and Schrodt (2011, 2014), Chadefaux (2014), Muchlinski et al. (2016), and Colaresi and Mahmood (2017) for examples of studies on the other end of the spectrum, which focus on methods that excel at prediction without attention to explanation. See Mearsheimer and Walt (2013) for a general critique of empirical research that is insufficiently theoretically grounded, and see Blair and Sambanis (2020) for a study that demonstrates the role of theory in improving predictive accuracy.

[18] In imposing the restriction, we are imposing a scope condition in which we can only explain crisis escalation dynamics among the set of states that had some level of cooperative or conflictual interaction—a minimal definition of political relevance. Specifically, we use the BBC Monitoring's Summary of World Broadcasts (1979–2019) from the Phoenix Event Data. Data can be accessed via https://doi.org/10.13012/B2IDB-0647142_V3. Accessed July 4, 2021.

[19] We are using ICB v14, which includes cases up to 2017.

inflict military, economic, and political costs against one another is constantly changing. Amid such changes, the ability for states to resolve their disputes peacefully depends in part on how much the distribution of power has become out of line with the distribution of benefits (Powell 1999; Reed et al. 2008; Carroll and Kenkel 2019). In principle, changes in relative material power could affect state calculations about whether to escalate or de-escalate a conflict (Gilpin 1981; Van Evera 1999; Copeland 2000; Lobell 2003). Some contend that dominant states in relative decline are more likely to escalate conflicts to push emerging challengers back down the power curve (Silverstone 2007; Levy and Thompson 2010). Powell (2006) models this as a commitment problem in which the declining state prefers war to negotiation because the rising state cannot commit to abide by a new bargained relationship. Organski's (1958) power transition theory maintains that dissatisfied rising states are more likely to escalate conflicts and even provoke war as a window of opportunity opens to alter the global or regional order in their favor. Alternatively, rising states might act cautiously in their interactions with more powerful states to avoid triggering a crisis and war (Van Evera 1999, 86). A declining state might also behave cautiously and seek de-escalation when interacting with a rising state to cultivate a stable relationship before it is eclipsed in power by the latter. In other words, given the option, a declining state would become likely to retrench from global commitments, concentrate its resources at home, and reduce the costs of leadership (MacDonald and Parent 2011).

We include the following suite of variables to capture relative rise and decline in terms of military capacity, foreign-policy preference similarity, trade dependence, and economic performance: (1) the level and change in military expenditure ratio (World Bank 2020)[20]; (2) the level and change in military personnel ratio (World Bank 2020); (3) the level and change in ideal point distance based on their voting records in the United Nations General Assembly (UNGA; Bailey, Strezhnev, and Voeten 2017)[21]; (4) dyadic trade dependence (level and change) using the Correlates of War (COW) Bilateral Trade Data (v 4.0)[22]; (5) and the level and change in GDP ratio from the Penn World Tables (PWT) 10.0 (Feenstra, Inklaar, and Timmer 2015).[23] These structural variables are commonly used as indicators of bargaining power in quantitative studies.[24]

A second set of variables pertains to what we term *regime entrenchment*, defined as the stability of the state and the degree to which it possesses the institutional capacity to govern, preserve order, and defend the regime from within (see Ripsman 2016). Variables that measure regime entrenchment include the level and change in the Polity composite democracy–autocracy index and the durability score from the Polity data (Marshall, Jaggers, and Gurr 2002).[25]

Building on the logic of Debs and Goemans (2010), regime entrenchment variables connected to the extent to which democratic governance is consolidated should do relatively well in explaining tipping-point behavior. Leaders in consolidated democracies face less severe punishment for making unpopular concessions and thus face fewer constraints as they pursue prudent de-escalation actions to avoid war. In contrast, leaders in non-democracies tend to face starker punishments for

---

[20]The ratio is calculated as the differences between the high and low values, weighted by the high value variable. A value near 1 indicates high asymmetry, while a value of 0.5 indicates high parity.

[21]See https://doi.org/10.7910/DVN/LEJUQZ or https://github.com/evoeten/United-Nations-General-Assembly-Votes-and-Ideal-Points. Accessed July 4, 2021.

[22]See http://www.correlatesofwar.org/data-sets/bilateral-trade. Accessed July 4, 2021. COW's Bilateral Trade Data are only updated to 2014, so we only use trade dependence variables in selected models.

[23]Like the military expenditure ratio, the GDP ratio is calculated as the differences between the high and the low values, weighted by the high-value variable. PWI 10.0 is available for download at https://www.rug.nl/ggdc/productivity/pwt/?lang=en. Accessed July 4, 2021.

[24]See Bennett and Stam (2004) for a comprehensive overview of the quantitative study of war.

[25]We use the latest Polity5 data. See http://www.systemicpeace.org/polityproject.html. Accessed July 4, 2021.

concessions and thus face greater incentives to unconditionally proceed down the path of crisis escalation once started. The leaders of states with low entrenchment are prone to find crisis escalation at tipping-point junctures net-beneficial as gambles to avoid concessions; they make their decisions based on the expectations of domestic punishment severity rather than solely for the purposes of deterring belligerent competitors or attempting to reach an efficient bargain.[26] This logic also leads to the expectation that regime entrenchment variables will condition the explanatory power of action-level variables. In the presence of low entrenchment, conflictual statements and actions should do relatively well in helping to anticipate whether the actors are locked into crisis escalation. In the presence of high entrenchment, the leaders will have more leeway to change course after earlier belligerent statements and actions, making action-level variables less informative.

A third set of variables that we consider to be influential in shaping the potential for crisis, but distinct from regime entrenchment, is the *stateness* of the actors, which we define as the domestic institutional centrality of a state and the ability to resist or overcome domestic political pressure (Evans 1997; Blanchard and Ripsman 2013). While also related to domestic political insecurities, this concept more closely captures the capability of leaders to fully direct the resources of the state in the midst of a potential international crisis. To measure the *stateness* of the states, we include the state fragility index for each state, which is an aggregation of eight component variables related to the effectiveness and legitimacy of each state regarding the security, political, economic, and social domains.[27] In addition, we include the number of violent non-state actors as coded in the Big Allied and Dangerous (BAAD) dataset (Asal, Rethemeyer, and Schoon 2019) in each country of the dyad, as these BAAD groups reflect the inability for the state to maintain a monopoly on the use of force.[28]

We also include a set of action-level variables, to bring in information about cooperative and conflictual events. The action-level variables come from the Phoenix event data project (Althaus et al. 2020), specifically the BBC Summary of World Broadcasts data, which is the only portion of the Phoenix project that covers our entire temporal domain. We use the Phoenix data to code counts of actions that fall into the categories of material conflict, material cooperation, verbal conflict, or verbal cooperation—so-called quad variables.[29] In addition to the monthly counts of the actions that fall into these categories, we included measures of the accumulation of actions by taking 12-month rolling sums of each "quad" category. The monthly counts reflect what the actors are doing in the immediate period prior to the observation, whereas the cumulative variables are more reflective of the general tendencies for the states to be more cooperative or conflictual toward each other in recent times. To avoid simultaneity bias, we lag each explanatory variable by one month.

To account for additional interdependencies among potential crisis actors, we include the physical geographic distance between each actor, measured as the natural log of their minimum distance (km) from the CShapes data (Weidmann, Kuse, and Gleditsch 2010). Finally, to account for differing baseline propensities for each actor to be involved in a crisis, we include the *degree centrality* of each actor in the

---

[26] These entrenchment variables will still have some limitations in predicting exactly when a crisis will tip at a high level of temporal precision because leaders will often have many choices in terms of the timing of the next move, even when leaders have a finite time to decide how to respond to a situation—the presence of "finite time" is a criterion for a situation to be in near crisis or crisis (Iakhnis and James 2021).

[27] For the State Fragility Index and Matrix, see https://www.systemicpeace.org/inscrdata.html. Accessed July 4, 2021.

[28] We use the latest version of the BAAD data, which covers the period from 1998 to 2017. These variables are only used in select models because they do not fully cover the temporal period of this study.

[29] For other uses of this "quad" typology, see Duval and Thompson (1980), Metternich, Minhas, and Ward (2017), Tellez and Roberts (2019), and Roberts and Tellez (2020).

international crisis network (Hafner-Burton, Kahler, and Montgomery 2009), which measures the extent to which actors are more likely to become targets of international crises (in-degree) or to target other actors (out-degree).[30] We create a set of longitudinal networks at the monthly level from 1995 to 2017 in which network ties reflect two actors engaging in international crisis. We then calculate the total degree centrality for each actor over the entire time period.

<div align="center">*Quantitative Results*</div>

We start with a random forest model (Liaw and Wiener 2002), which is a type of machine learning "classification tree" model that excels at comparing how much information is added by including each variable. Classification tree models divide the multidimensional space created by a set of variables (in this case, the explanatory variables discussed in the previous section) attached to each observation into separate regions. Each of these regions is represented by a terminal node of the tree. The objective is for each region to maximize the proportion of observations belonging to a single category (in case of a binary outcome like ours, there are two categories: crisis onset or no crisis onset). A classification tree begins with all observations at an origin node and is generated by repeatedly dividing between two branches representing differences in the values of an explanatory variable. The random forest model builds upon this idea of a classification tree by randomly choosing between a subset of variables for each division and then averaging across all classification trees generated (Breiman 2001). Estimation of random forest models is computationally intensive, so we aggregated the "quad" action-level variables into cooperative and conflictual events.

Figure 1 presents the ranked order of the variables that have the most explanatory power in predicting the onset of an ICB crisis.[31] We see that when compared to the top three variables, all other variables have poor explanatory power. The top three variables, moreover, are all the most direct indicators of whether permissive conditions for crisis escalation are present. High-degree centrality by either actor, as well as the accumulation of conflictual events in the dyad, does well to help anticipate the onset of an international crisis.

To further evaluate the explanatory power of the variables in the random forest model, we generate a table that compares the predicted to the actual values.[32] We first identify a "best threshold" of a probability of a predicted crisis above which we classify a case as being a likely crisis. We use Youden's J statistic (Youden 1950) to identify the threshold that maximizes both sensitivity and specificity.[33] Once we identify this "threshold" from the training sets, we use it to classify the data and generate the confusion matrix for our in-sample forecasting, presented in table 1.

Table 1 demonstrates that the random forest model does well to avoid false negatives, but the number of false positives is higher than the true positives. We see here that the model does well to identify a set of cases that are not at risk for crisis escalation—no cases escalated when the probability of escalation was below the threshold. The model, however, is a poor predictor of crisis escalation among the cases that are at the greatest risk. The results are consistent with what we would

---

[30] We sum the in-degree and out-degree centrality measures, that is, total degree centrality.

[31] The relative importance is based on the total decrease in node impurities from splitting on the variable, averaged over all trees.

[32] This is sometimes referred to as a "confusion matrix." The predictions are in-sample. In models not shown, out-of-sample prediction performed by using data from 1991–2007 as the training set and data from 2008–2015 as the test set performs even worse, further demonstrating the challenges of accurately anticipating crises.

[33] We use the information from the Receiver Operating Characteristic (ROC) curve to produce an optimal cutoff as the best threshold that maximizes the distance to the identity (diagonal) line of the curve. This procedure is performed using the R package pROC. See Robin et al. (2011).
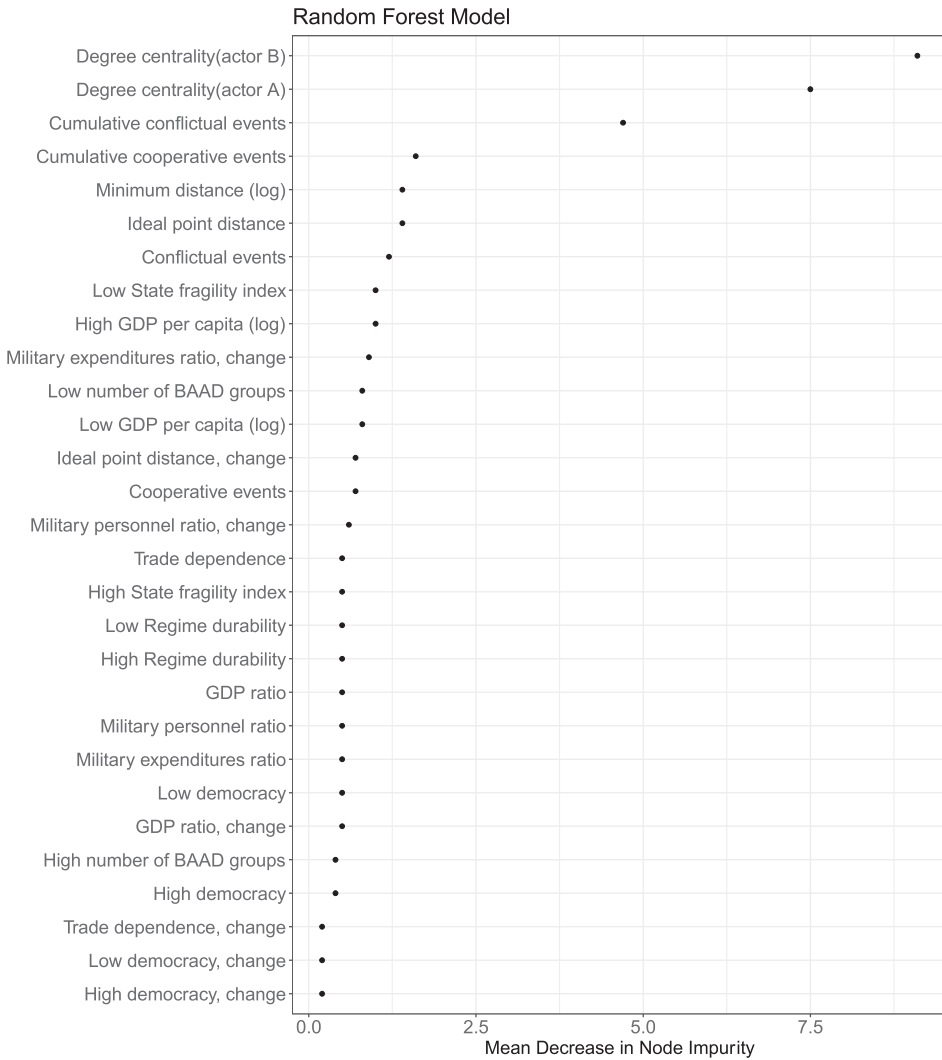
Figure 1. Random forest model of crisis onset.

Table 1. Predicted versus actual crises for random forest model

		Predicted values	
		Negative (0)	Positive (1)
Actual values	Negative (0)	1,090,305	45
	Positive (1)	0	42

expect from variables that capture the permissive conditions for crisis but not the immediate causes of crisis.

The findings here confirm that quantitative models can do well to understand when the permissive conditions for crisis are present, but they are less well suited to explain the immediate causes of crisis escalation. To get a clearer sense of how well variables can explain tipping-point behavior within the set of cases that are prone to escalate, we estimate an SPDM (Schmidt and Witte 1989; Beger et al. 2017). The

SPDM estimates two models—one for whether a given observation is at risk for ex-
periencing crisis onset and another for the time until onset among the observations
at risk. That is, the SPDM is designed to account for the potential for many obser-
vations to not be at risk for experiencing the event in question and thus can help
separate permissive conditions from immediate causes of conflict (Beger, Dorff, and
Ward 2016).

Figure 2 presents the results for the two equations that the SPDM estimated.[34]
The set of coefficients on the left shows the relationships with the risk of ever ex-
periencing a crisis, and the coefficients on the right show the relationships with the
time until a crisis does occur conditional on the potential to be at risk. The figure
reveals four striking patterns. First, the degree centrality variables dominate as ex-
planatory variables for being at risk for crisis occurrence, which corroborates the
patterns observed in the random forest model. Second, in focusing on the time-
to-onset equation, the regime entrenchment and the *stateness* variables tend to do
well to explain the timing of a crisis, conditional on being at risk for crisis. Consis-
tent with expectations, domestic pressures on leaders to avoid concessions can push
them further toward crisis escalation once on the path. Third, the structural vari-
ables related to rise and decline in power tend to do poorly in explaining the onset
of crisis, with the exception of the low GDP per capita and GDP ratio variables.
Fourth, many of the action-level variables also do relatively poorly in explaining
the timing of crisis escalation, with the exception of material conflict between the
adversaries. These latter two findings are consistent with expectations—structural
variables related to state power dynamics, as well as action-level variables do not do
well to explain tipping-point behavior.

We not only expect the regime entrenchment and stateness variables to have the
most explanatory power in the propensity for crisis escalation among the set of at-
risk cases but also expect these domestic politics variables to condition the extent
to which the action-level variables inform tipping-point behavior. We thus run an
additional SPDM model that is subset on low regime entrenchment—measured as
whether a state in the dyad has a low Polity score—to see how the values of the other
variables change based on whether the states involved are relatively authoritarian or
democratic. The actions of leaders with low regime entrenchment are expected to
do relatively well in anticipating the escalation of crises because their choices are
more likely made for the benefit of domestic audiences and can potentially lock
in a leader to a course of action, rather than solely being the product of strategic
calculations meant to deter an opponent or find an efficient bargaining outcome.

Figure 3 shows the results for the dyads in which the low democracy score is below
zero. We now observe that all of the action-level variables are statistically significant
at explaining the time until crisis onset. Moreover, the relative sizes of the substan-
tive effects are large. When actors with low regime entrenchment mobilize and pos-
ture, they become much more likely to end up in crisis. Their observed conflictual
and cooperative actions thus are more informative regarding whether a crisis is
about to escalate than the actions of states with high regime entrenchment. These
findings are consistent with an understanding from Debs and Goemans (2010) of
non-democratic leaders facing situations in which backing down is more costly to
the leader than going forward with crisis escalation.

Strengths of Qualitative Approaches

Qualitative methods offer complementary strengths to their quantitative counter-
parts. With careful case studies, researchers can drill down and access much more
of the data—even those that may not be readily available for other cases—in or-
der to get a comprehensive and specific picture of the interaction, rather than the

---

[34] The coefficient plots are rescaled so that the standard errors are constant.
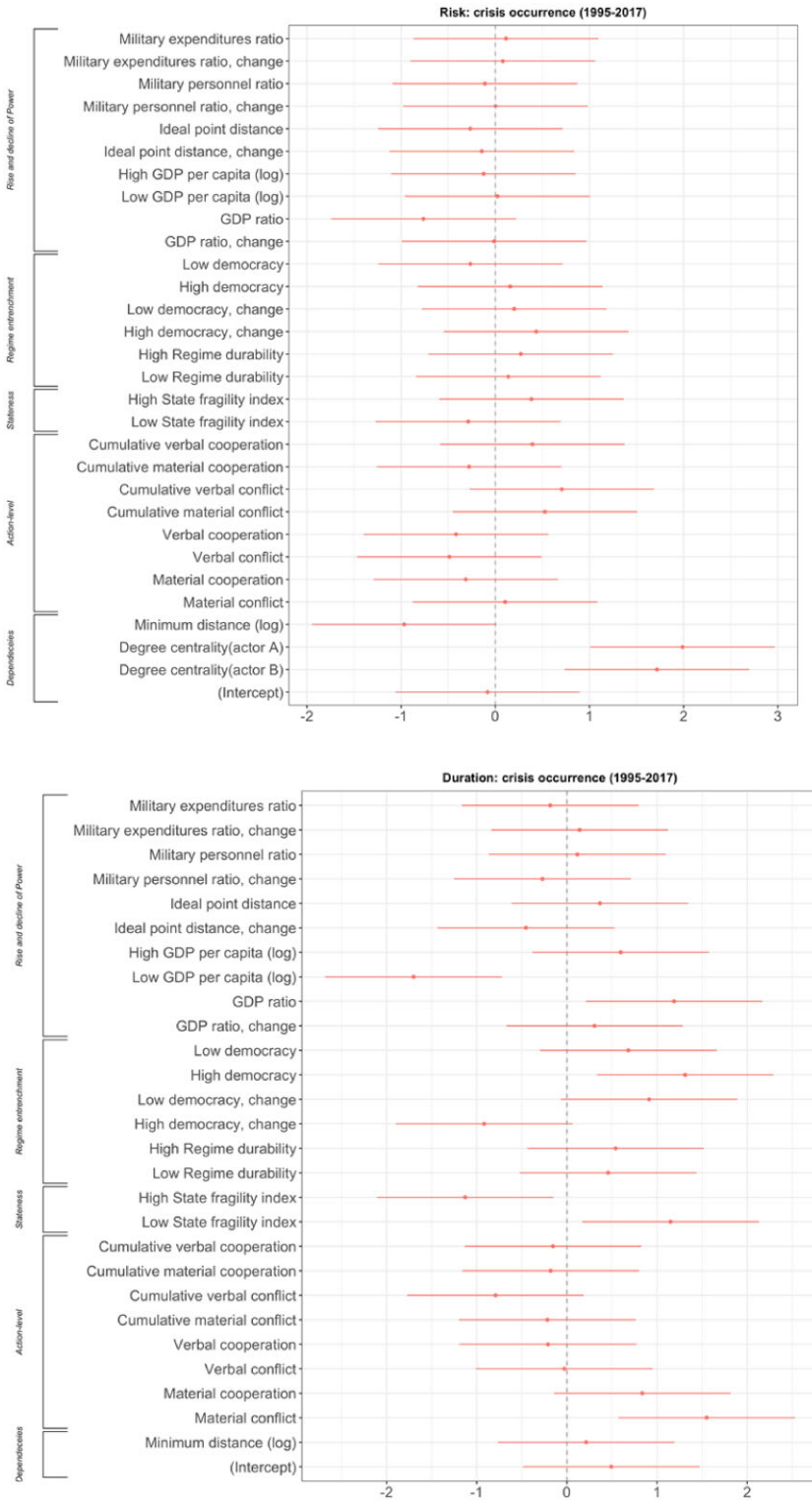
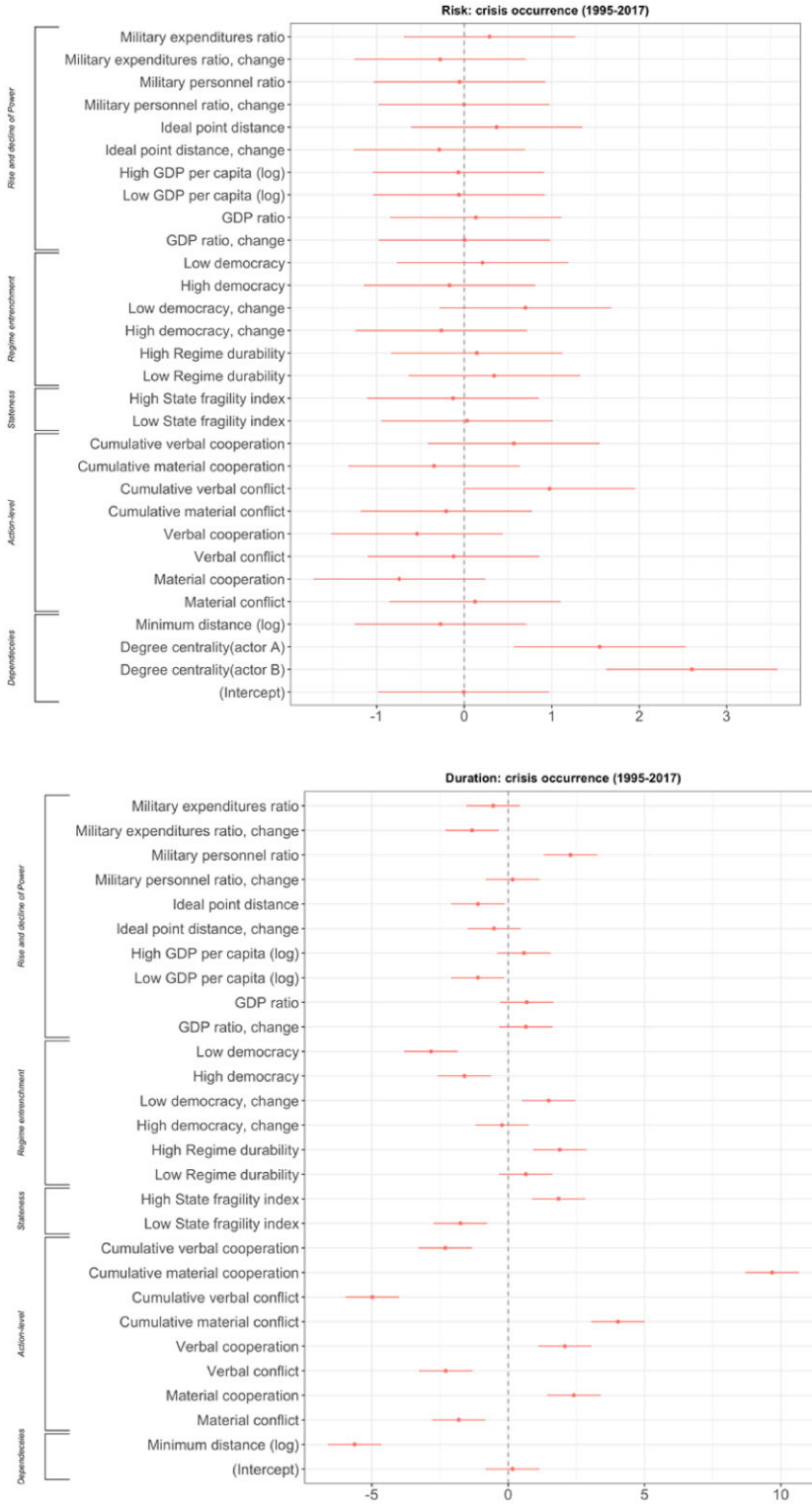**Figure 2.** SPDM models of crisis risk and incidence, with 95 percent confidence intervals.

Figure 3. SPDM models, low Polity sample.

superficial representation that aggregated data points provide. Yet, the intensive study of individual cases takes time and resources that make it impossible to address a sufficient number of cases to generalize meaningfully, which is the core mission of social science research. While it is possible to mitigate this problem somewhat with a careful case selection strategy—such as selecting potential crucial cases, hard tests of the argument, etc.—absent a sufficiently large number of observations, it is impossible to determine whether the underlying dynamic observed is idiosyncratic or a relationship that operates across the universe of cases (George and Bennett 2005; Gerring 2007; Weller and Barnes 2014).

We discuss three particular advantages that qualitative, historical methods bring to the explanation of tipping-point behavior: (1) the ability to explore the entirety of the episode using information and context that would be overlooked by statistical analysis; (2) the potential to find "smoking gun" evidence that links cause and effect; and (3) the sensitivity to leader perceptions.

Information Advantages

With historical case study methods, the researcher gains the ability to discover the presence or absence of catalytic factors that cannot be coded systematically across cases but that can be identified using archival and discursive tools. By exploring the private papers of the key decision-makers in multiple states, examining Cabinet minutes (which can be done for older cases) or by interviewing the participants in the crisis (which can be done for more recent cases), we can go beyond the variables that we have coded and engage in a more open-ended analysis about the factors that led specific near crises to tip into crises (e.g., George and Bennett 2005). This might lead to the discovery of causal paths and variables not easily included or measured in quantitative datasets, but which could help anticipate similar future events by comparing real-time patterns in leader statements and observed behavior with historical cases of similar patterns in which the underlying thought processes beneath the statements and behavior are well known.

Related to the point above about incentives to avoid inefficiencies, qualitative evidence can play a helpful role, similar to insider trading information or other forms of information arbitrage. A case study that considers what leaders would have known and not known at crucial junctures can elucidate information that might not have been known to the adversaries or other relevant third parties at the time. It can also shed light on the priorities of the decision-makers. Consequently, it can provide useful information about how leaders interacted with and responded to the structural variables discussed above and what may lead to escalation within a given set of structural conditions.

Discovery of "Smoking Gun" Evidence

Because of the wealth of information, including the access to government documents, private paper collections, memoirs, and even interviews with the decision-makers, which qualitative case studies can utilize, they have the potential to discover "smoking gun" evidence that explains the exact reasons why a particular decision was made. This allows researchers not merely to correlate variables with outcomes or sequence hypothesized cause with effect but rather to demonstrate—at least in an individual case—what caused the outcome in question (George and Bennett 2005). When researching what tips near crises into crises, we can use qualitative methods to go beyond the permissive conditions to determine what exactly was the catalyst for escalation, at least in individual cases. After all, data can describe environmental factors within which decisions are made; firsthand reports can explain motives.

The Role of Perception

Qualitative research can elucidate how the actors in question actually understood the situation they were in, rather than merely the objective data points that are quantifiable in the dataset. If actors misperceive the relative balance of power, the intentions of other actors, or likely domestic consequences of acting in a particular way, and if they act based on these misperceptions, it would be inaccurate to ascribe causal importance to objective environmental conditions (Ripsman, Taliaferro, and Lobell 2016, 131–37).[35] This may help explain why environmental conditions alone are poor predictors of escalation to crisis.

Discovering the development of actors' perceptions and interpretations of their adversaries' perceptions is particularly important in studying near-crises and crises, since by their very definition crises and near crises include perceptual elements in their determination. As mentioned above, the first defining feature of an ICB crisis is the perception of a risk to basic values. The second defining characteristic of a crisis is the perception of a defined time frame by one or more actors, and the third defining feature of a crisis is the perception of a heightened risk of military escalation by one or more actors. In the case of a near crisis, actors perceive a shortened time horizon and a threat to basic values but no heightened risk of military escalation (Iakhnis and James 2021). Therefore, the ability to explore actors' perceptions is crucial to determining crisis dynamics and tipping points.

A Tipping-Point Example: The Third Taiwan Straits Crisis

We can illustrate the importance of qualitative analysis for determining tipping-point trajectory with the Third Taiwan Straits Crisis, 1995–1996.[36] We present this case because our qualitative analysis reveals that the crisis trigger date identified by the ICB dataset does not well capture the actual moment that the case tipped into crisis. In this case, the near crisis began with PRC leader Jiang Zemin's major speech on Taiwan on January 30, 1995, in honor of the Chinese New Year. Jiang offered a more extensive economic relationship with Taiwan, as well as additional concessions, provided Taipei quickly negotiated unification with the mainland on the basis of the One China policy. At the same time, however, he threatened that an unacceptable delay in reaching unification could force Beijing to consider the use of force, which he would not rule out (Copper 1999, 41–42; Jiang 1999). To reinforce this threat, from February through April 1995, the People's Liberation Army (PLA) engaged in military exercises and force redeployments that escalated pressure on Taiwan, while Jiang and his foreign ministry escalated the political rhetoric (Wu 1995; Zhu and Lu 1995; Copper 1999, 41).

This situation did not tip into crisis, however, until after Taiwanese leader Lee Teng-hui traveled to the United States on a tourist visa on June 7 to participate in a Cornell University reunion. The PRC leadership felt that this was an unacceptable challenge to the One China policy, and they were angry with both Lee for his brazen disrespect and Washington for violating previous commitments. Between July 1995 and March 1996, the PRC responded with several rounds of missile tests and live-fire military exercises in the waters adjacent to Taiwan.

Evaluating the events data alone, the tipping point appears to be Lee's June 7 visit. Indeed, that is the ICB project's conclusion: the ICB crisis began on June 7 and ended in March 1996.[37] Our case analysis, however, revealed that this was not in fact the tipping point from near crisis into crisis. To be sure, Beijing was very unhappy with the visit, but it was not a forgone conclusion that the PRC would *escalate* vis-à-vis

---

[35] See also Yarhi-Milo (2014).

[36] The crisis number in the ICB project is 415; for a case summary of this crisis, see http://www.icb.umd.edu/dataviewer/?crisno=415. Accessed July 4, 2021.

[37] For the ICB case summary, see http://www.icb.umd.edu/dataviewer/?crisno=415. Accessed July 4, 2021.

Taiwan and the United States as a result. In fact, Lee's office had a secret channel through which it communicated to Jiang, through which they informed Jiang in advance that Lee would seek to travel to the United States in June. Jiang's representative acknowledged this information and did not warn of any consequences that would follow from an unofficial visit of this sort. The Taiwanese leadership, therefore, did not believe that there would be any Chinese retaliation.[38] Moreover, it took a full month before Beijing announced that the PLA would commence missile tests and exercises near Taiwan later in July (Garver 1997, 75–76; Copper 1999, 42). If the Lee visit was the tipping point, why did it take so long to announce and initiate escalation?

The answer lies in Chinese domestic politics. The CCP was due to hold its fifteenth National Congress in 1997, heightening the sense of urgency for Jiang to consolidate political support. Jiang was considered an outsider in Beijing. Having served as the mayor of Shanghai, which was peripheral to the seat of Chinese power, Jiang was viewed as a "provincial leader" in Beijing's elitist circles. Moreover, the PLA, a crucial power broker in the PRC, viewed Jiang as "a pure civilian" with no military experience. Especially with regard to Taiwan, the PLA believed that Jiang and his foreign minister Qian Qichen were too weak to defend Chinese interests (see Garver 1997, 61; Lam 1999, 173). Jiang was consequently under increasing pressure to satisfy the military, which was much more bellicose than PRC political leaders on the Taiwan issue, and believed that Qian "had been played for a patsy" by Washington and Taipei (Scobell 2000, 235).

This conflict came to a head with the Lee visit to Cornell. Initially, the moderates prevailed and the Chinese reaction was limited to a public protest. At a contentious meeting of the Leading Group on Taiwan Affairs in mid-June, however, the PLA pressured Jiang to take a harder line, using military means (Garver 1997, 61–62, 73). In this context, if Jiang was going to solidify his position as Communist Party Leader, he needed to take a harder line against Taiwan to satisfy the PLA. This explains why Chinese escalation did not commence until many weeks after Lee's visit to the United States. The catalyst, therefore, was Lee's visit, but the tipping point was the intense PLA pressure and criticism later in June.

As the Third Taiwan Straits crisis demonstrates, analysts would miss a key explanation that would help anticipate why the Lee visit triggered the crisis for China if they only had the type of information that would be typically included in quantitative studies of crisis escalation. Without an understanding of the specific political pressures that Jiang faced, we would not be able to have anticipated with high accuracy that the Lee visit at that moment in time would have triggered a crisis for China.

This is not to say that quantitative models would have been useless in trying to anticipate the crisis. A number of variables in our quantitative analysis above, such as those related to changes in relative power, the cumulative amount of conflict over the Taiwan problem, and the increase in verbal conflictual rhetoric and military exercises in early 1995 would certainly have placed the probability of crisis escalation in June of 1995 at above zero. The case, however, also suggests that a more leader-specific entrenchment variable would add additional information to the quantitative model. Relevant for this case would have been information on Jiang Zemin's recent accession to power and the shadow of the forthcoming fifteenth National Congress. When considering the concept of regime entrenchment, this case points to ways in which the above quantitative models could have been improved, for example, by using a measure of leader tenure rather than regime durability.

Nevertheless, even with better quantitative measures, we still would miss the explanation of why this tipped into a crisis without the full picture of the specific

---

[38] Interview with Su Chi, former Deputy Secretary General of the President's Office, 1995–2000, March 8, 2019. See Su (2008).

cleavages within the PLA that a careful qualitative analysis can provide—the considerations of the decision-maker incentives and perceptions that are at a level that is difficult to systematically code across cases.

Conclusions

Explaining the occurrence of crisis escalation based on information in structural variables and action-level variables is hard for a number of different reasons. Some of the reasons stand to be reduced with further advances in machine learning model estimation and text-as-data processing. With methodological advances, more cases can be coded and coded with higher levels of accuracy. Empirical models *will* continue to improve, without question.

At the same time, we present reasons why we should expect quantitative models never to approach perfection. The actions of strategic actors will always have a strong stochastic component. Furthermore, as the Taiwan Straits Crisis case shows, contextual knowledge is often important to explain whether and when a particular high probability crisis will escalate. However, this does not mean that actor decisions need to be treated as idiosyncratic. With a combination of qualitative and quantitative approaches, and with attention not only to structural variables but also to variables related to regime entrenchment and state legitimacy, researchers can explain much of when tipping-point moments occur and why they tipped one way or the other. Depending on the availability of information to the analyst regarding leader statements and behavior, a combined approach can help both in real-time anticipation of international crises and in the testing of competing hypotheses that expect different courses of action in a given situation. The advancement of knowledge about crisis escalation will depend on advances in both quantitative and qualitative approaches.

This study's implications extend beyond academic research to intelligence analysis and foreign policy decision-making. The challenges of anticipating crisis escalation also plague the IC (Dahl 2017; Zegart and Morell 2019) and proclamations by foreign policy experts (Drezner 2021). While skepticism of quantitative methods already exists within the IC (Bang 2016), quantitative models are improving in making use of the increasing amounts of information available to analysts (Dhami et al. 2015; Friedman et al. 2018; Lustick and Tetlock 2021). At the same time, the use of case study methods based on state-of-the-art social scientific methodology can greatly improve the accuracy of intelligence assessments that would otherwise be based on mere anecdotes (Dahl 2017). This study points to the importance of intelligence assessments to not place too much emphasis on the meaning of structural and action-level variables that potentially tell us little about how the adversary will respond to a course of action, while also emphasizing the importance of variables related to regime entrenchment and stateness that capture the incentives on leaders to take certain courses of action independent of the potential success of the actions.

Acknowledgments

the American Political Science Association and at the 2021 Tsinghua International Conference on Computational Social Science.

Funding

References

ALTHAUS, SCOTT, JOSEPH BAJJALIEH, JOHN F. CARTER, BUDDY PEYTON, AND DAN A. SHALMON. "Cline Center Historical Phoenix Event Data. v.1.3.0." Distributed by Cline Center for Advanced Social Research, University of Illinois at Urbana-Champaign, May 4, 2020. Accessed July 4, 2021. Accessed July 4, 2021. https://doi.org/10.13012/B2IDB-0647142_V3.

ASAL, VICTOR H., R. KARL RETHEMEYER, AND ERIC W. SCHOON. 2019. "Crime, Conflict, and the Legitimacy Trade-off: Explaining Variation in Insurgents' Participation in Crime." *The Journal of Politics* 81 (2): 399–410.

BAILEY, MICHAEL A., ANTON STREZHNEV, AND ERIK VOETEN. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61 (2): 430–56.

BANG, MARTIN. 2016. "Pitfalls in Military Quantitative Intelligence Analysis: Incident Reporting in a Low Intensity Conflict." *Intelligence and National Security* 31 (1): 49–73.

BEGER, ANDREAS, CASSY L. DORFF, AND MICHAEL D. WARD. 2016. "Irregular Leadership Changes in 2014: Forecasts Using Ensemble, Split-Population Duration Models." *International Journal of Forecasting* 32 (1): 98–111.

BEGER, ANDREAS, DANIEL W. HILL, JR., NILS W. METTERNICH, SHAHRYAR MINHAS, AND MICHAEL D. WARD. 2017. "Splitting It Up: The Spduration Split-Population Duration Regression Package for Time-Varying Covariates." *The R Journal* 9 (2): 474–86.

BEGER, ANDREAS, RICHARD K. MORGAN, AND MICHAEL D. WARD. 2021. "Reassessing the Role of Theory and Machine Learning in Forecasting Civil Conflict." *Journal of Conflict Resolution* 65 (7–8): 1405–26.

BENNETT, D. SCOTT, AND ALLAN C. STAM, III 2004. *The Behavioral Origins of War.* Ann Arbor, MI: University of Michigan Press.

BETTS, RICHARD K. 2010. *Surprise Attack: Lessons for Defense Planning.* Washington, DC: Brookings Institution Press.

BLAIR, ROBERT A., AND NICHOLAS SAMBANIS. 2020. "Forecasting Civil Wars: Theory and Structure in an Age of 'Big Data' and Machine Learning." *Journal of Conflict Resolution* 64 (10): 1885–915.

———. 2021. "Is Theory Useful for Conflict Prediction? A Response to Beger, Morgan, and Ward." *Journal of Conflict Resolution* 65 (7–8): 1427–53.

BLANCHARD, JEAN-MARC F., AND NORRIN M. RIPSMAN. 2013. *Economic Statecraft and Foreign Policy: Sanctions, Incentives and Target State Calculations.* New York: Routledge.

BOSCHEE, ELIZABETH, JENNIFER LAUTENSCHLAGER, SEAN O'BRIEN, STEVE SHELLMAN, AND JAMES STARZ. 2018. "ICEWS Weekly Event Data." Harvard Dataverse, V263. Accessed July 4, 2021. https://doi.org/10.7910/DVN/QI2T9A.

BOWLSBY, DREW, ERICA CHENOWETH, CULLEN HENDRIX, AND JONATHAN D. MOYER. 2020. "The Future is a Moving Target: Predicting Political Instability." *British Journal of Political Science* 50 (4): 1405–17.

BRANDT, PATRICK T., JOHN R. FREEMAN, AND PHILIP A. SCHRODT. 2011. "Real Time, Time Series Forecasting of Inter- and Intra-State Political Conflict." *Conflict Management and Peace Science* 28 (1): 41–64.

———. 2014. "Evaluating Forecasts of Political Conflict Dynamics." *International Journal of Forecasting* 30 (4): 944–62.

BRECHER, MICHAEL, AND JONATHAN WILKENFELD. 2000. *A Study of Crisis.* Ann Arbor, MI: University of Michigan Press.

BRECHER, MICHAEL, JONATHAN WILKENFELD, KYLE BEARDSLEY, PATRICK JAMES, AND DAVID QUINN. 2020. International Crisis Behavior Data Codebook, Version 13. Accessed July 4, 2021. http://sites.duke.edu/icbdata/data-collections/.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Bueno de Mesquita, B. 2011. "A New Model for Predicting Policy Choices: Preliminary Tests." *Conflict Management and Peace Science* 28 (1): 65–87.

Bunge, Mario. 1996. *Finding Philosophy in Social Science.* New Haven, CT: Yale University Press.

Carroll, Robert J., and Brenton Kenkel. 2019. "Prediction, Proxies, and Power." *American Journal of Political Science* 63 (3): 577–93.

Carrubba, Clifford J., Amy Yuen, and Christopher Zorn. 2007. "In Defense of Comparative Statics: Specifying Empirical Tests of Models of Strategic Interaction." *Political Analysis* 15 (4): 465–82.

Cederman, Lars-Erik, and Nils B. Weidmann. 2017. "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science* 355 (6324): 474–76.

Chadefaux, Thomas. 2014. "Early Warning Signals for War in the News." *Journal of Peace Research* 51 (1): 5–18.

Chenoweth, Erica, and Jay Ulfelder. 2017. "Can Structural Conditions Explain the Onset of Nonviolent Uprisings?" *Journal of Conflict Resolution* 61 (2): 298–324.

Chiba, Daina, and Kristian Skrede Gleditsch. 2017. "The Shape of Things to Come? Expanding the Inequality and Grievance Model for Civil War Forecasts with Event Data." *Journal of Peace Research* 54 (2): 275–97.

Colaresi, Michael, and Zuhaib Mahmood. 2017. "Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting." *Journal of Peace Research* 54 (2): 193–214.

Copeland, Dale C. 2000. *The Origins of Major War.* Ithaca, NY: Cornell University Press.

Copper, John F. 1999. "The Origins of Conflict across the Taiwan Strait: The Problem of Differences in Perception." In *Across the Taiwan Strait: Mainland China, Taiwan, and the 1995–96 Crisis*, edited by Suisheng Zhao, 41–42. New York: Routledge.

Croco, Sarah E. 2011. "The Decider's Dilemma: Leader Culpability, War Outcomes, and Domestic Punishment." *American Political Science Review* 105 (3): 457–77.

Dahl, Erik J. 2017. "Getting beyond Analysis by Anecdote: Improving Intelligence Analysis through the Use of Case Studies." *Intelligence and National Security* 32 (5): 563–78.

Debs, Alexandre, and Hein E. Goemans. 2010. "Regime Type, the Fate of Leaders, and War." *American Political Science Review* 104 (3): 430–45.

Dhami, Mandeep K., David R. Mandel, Barbara A. Mellers, and Philip E. Tetlock. 2015. "Improving Intelligence Analysis with Decision Science." *Perspectives on Psychological Science* 10 (6): 753–57.

Drezner, Daniel W. 2003. "The Hidden Hand of Economic Coercion." *International Organization* 57 (3): 643–59.

———. 2021. "The Perils of Predictions." *Foreign Policy* 239: 28–33.

Duval, Robert D., and William R. Thompson. 1980. "Reconsidering the Aggregate Relationship between Size, Economic Development, and Some Types of Foreign Policy Behavior." *American Journal of Political Science* 24 (3): 511–25.

Evans, Peter B. 1997. "The Eclipse of the State? Reflections on Stateness in an Era of Globalization." *World Politics* 50 (1): 62–87.

Fearon, James D. 1994. "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38 (2): 236–69.

———. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.

———. 2002. "Selection Effects and Deterrence." *International Interactions* 28 (1): 5–29.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82.

Friedman, Jeffrey A., Joshua D. Baker, Barbara A. Mellers, Philip E. Tetlock, and Richard Zeckhauser. 2018. "The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament." *International Studies Quarterly* 62 (2): 410–22.

Gartner, Scott Sigmund, and Jacob Bercovitch. 2006. "Overcoming Obstacles to Peace: The Contribution of Mediation to Short-Lived Conflict Settlements." *International Studies Quarterly* 50 (4): 819–40.

Gartzke, Erik. 1999. "War is in the Error Term." *International Organization* 53 (3): 567–87.

Garver, John W. 1997. *Face Off: China, the United States, and Taiwan's Democratization.* Seattle, WA: University of Washington Press.

George, Alexander, and Andrew Bennett. 2005. *Case Studies and Theoretical Development in the Social Sciences.* Cambridge, MA: MIT Press.

Gerring, John. 2007. "Is There a (Viable) Crucial-Case Method?" *Comparative Political Studies* 40 (3): 231–53.

Gilpin, Robert. 1981. *War and Change in World Politics.* Princeton, NJ: Princeton University Press.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46 (5): 712–24.

Gleditsch, Kristian Skrede, Kyle Beardsley, and Sara M.T. Polo. 2010. "Issues in Data Collection: International Conflict." Oxford Research Encyclopedia of International Studies. Accessed July 4, 2021. https://doi.org/10.1093/acrefore/9780190846626.013.100.

Gleditsch, Kristian Skrede, and Michael D. Ward. 2013. "Forecasting is Difficult, Especially about the Future: Using Contentious Issues to Forecast Interstate Disputes." *Journal of Peace Research* 50 (1): 17–31.

Goemans, H.E. 2010. *War and Punishment.* Princeton, NJ: Princeton University Press.

Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, and Jay Ulfelder et al. 2010. "A Global Model for Forecasting Political Instability." *American Journal of Political Science* 54 (1): 190–208.

Hafner-Burton, Emilie M., Miles Kahler, and Alexander H. Montgomery. 2009. "Network Analysis for International Relations." *International Organization* 63 (3): 559–92.

Hegre, Håvard, Håvard Mokleiv Nygård, and Peder Landsverk. 2021. "Can We Predict Armed Conflict? How the First 9 Years of Published Forecasts Stand Up to Reality." *International Studies Quarterly* 65 (3): 660–68.

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand, and Henrik Urdal. 2013. "Predicting Armed Conflict, 2010–2050." *International Studies Quarterly* 57 (2): 250–70.

Hegre, Håvard, Marie Allanssona, Matthias Basedaua, Michael Colaresi, Mihai Croicua, Hanne Fjeldea, and Frederick Hoylesa et al. 2019. "Views: A Political Violence Early-Warning System." *Journal of Peace Research* 56 (2): 155–74.

Hollenbach, Florian, Iavor Bojinov, Shahryar Minhas, Nils W. Metternich, Michael D. Ward, and Alexander Volfovsky. 2021. "Multiple Imputation Using Gaussian Copulas." *Sociological Methods & Research* 50 (3): 1259–83.

Holsti, Ole R. 1972. *Crisis Escalation War.* Montreal: McGill-Queen's University Press.

Horowitz, Michael C., Alan C. Stam, and Cali M. Ellis. 2015. *Why Leaders Fight.* Cambridge: Cambridge University Press.

Iakhnis, Evgeniia, and Patrick James. 2021. "Near Crises in World Politics: A New Dataset." *Conflict Management and Peace Science* 38 (2): 224–43.

James, Patrick. 2019. "Systemist International Relations." *International Studies Quarterly* 63 (4): 781–804.

Jenke, Libby, and Christopher Gelpi. 2017. "Theme and Variations: Historical Contingencies in the Causal Model of Interstate Conflict." *Journal of Conflict Resolution* 61 (10): 2262–84.

Jiang, Zemin. 1999. "Document 3: General Secretary Jiang Zemin, 'Eight-Point Statement,' 30 January 1995." *Asian Affairs: An American Review* 26 (2): 100–101.

Kertzer, Joshua D., and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60 (1): 234–49.

Lam, Willy Wo-Lap. 1999. *The Era of Jiang Zemin.* Singapore: Simon and Schuster.

Lemke, Douglas, and William Reed. 2001. "The Relevance of Politically Relevant Dyads." *Journal of Conflict Resolution* 45 (1): 126–44.

Levy, Jack S., and William R. Thompson. 2010. "Balancing on Land and at Sea: Do States Ally against the Leading Great Power?" *International Security* 35 (1): 7–43.

Lewis, Jeffrey B., and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11 (4): 345–67.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by Randomforest." *R News* 2 (3): 18–22.

Lobell, Steven E. 2003. *The Challenge of Hegemony: Grand Strategy, Trade, and Domestic Politics.* Ann Arbor, MI: The University of Michigan Press.

Lustick, Ian S., and Philip E. Tetlock. 2021. "The Simulation Manifesto: The Limits of Brute-Force Empiricism in Geopolitical Forecasting." *Futures & Foresight Science* 3 (2): 1–22.

Luttwak, Edward. 2001. *Strategy: The Logic of War and Peace.* Cambridge, MA: Harvard University Press.

MacDonald, Paul K., and Joseph M. Parent. 2011. "Graceful Decline? The Surprising Success of Great Power Retrenchment." *International Security* 35 (4): 7–44.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2002. *Polity IV Project.* Accessed July 4, 2021. http://www.systemicpeace.org/inscrdata.html.

Mearsheimer, John J., and Stephen M. Walt. 2013. "Leaving Theory Behind: Why Simplistic Hypothesis Testing is Bad for International Relations." *European Journal of International Relations* 19 (3): 427–57.

Metternich, Nils W., Shahryar Minhas, and Michael D. Ward. 2017. "Firewall? Or Wall on Fire? A Unified Framework of Conflict Contagion and the Role of Ethnic Exclusion." *Journal of Conflict Resolution* 61 (6): 1151–73.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher. 2016. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24 (1): 87–103.

Oneal, John R., and Bruce M. Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41 (2): 267–93.

Organski, A.F.K. 1958. *World Politics.* New York: Alfred A. Knopf.

Pevehouse, Jon C.W., Timothy Nordstrom, Roseanne W. McManus, and Anne Spencer Jamison. 2020. "Tracking Organizations in the World: The Correlates of War IGO Version 3.0 Datasets." *Journal of Peace Research* 57 (3): 492–503.

Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics.* Princeton, NJ: Princeton University Press.

———. 2006. "War as a Commitment Problem." *International Organization* 60 (1): 169–203.

Reed, William, David H. Clark, Timothy Nordstrom, and Wonjae Hwang. 2008. "War, Power, and Bargaining." *The Journal of Politics* 70 (4): 1203–16.

Ripsman, Norrin M. 2016. *Peacemaking from Above, Peace from Below: Ending Conflict between Regional Rivals.* Ithaca, NY: Cornell University Press.

———. 2022. "Historical Methods." In *Handbook of Research Methods in International Relations*, edited by Joseph Huddleston, Thomas Jamieson and Patrick James, 74–88. Cheltenham: Edward Elgar Publishing.

Ripsman, Norrin M., Jeffrey W. Taliaferro, and Steven E. Lobell. 2019. *Neoclassical Realist Theory of International Politics.* New York: Oxford University Press.

Roberts, Jordan, and Juan Tellez. 2020. "Freedom House's Scarlet Letter: Negative Assessments and Verbal Conflict." In *The Power of Global Performance Indicators*, edited by Judith Kelley and Beth Simmons, 242–68. New York: Cambridge University Press.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. 2011. "Proc: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves." *BMC Bioinformatics* 12 (77): 1–8.

Samuelson, Paul A. 1965. "Proof That Properly Anticipated Prices Fluctuate Randomly." *Management Review* 6 (2): 41–49.

Schmidt, Peter, and Ann Dryden Witte. 1989. "Predicting Criminal Recidivism Using 'Split Population' Survival Time Models." *Journal of Econometrics* 40 (1): 141–59.

Schneider, Gerald, Nils Petter Gleditsch, and Sabine Carey. 2011. "Forecasting in International Relations: One Quest, Three Approaches." *Conflict Management and Peace Science* 28 (1): 5–14.

Scobell, Andrew. 2000. "Show of Force: Chinese Soldiers, Statesmen, and the 1995–1996 Taiwan Strait Crisis." *Political Science Quarterly* 115 (2): 227–46.

Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289–310.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93 (2): 279–97.

Silverstone, Scott. 2007. *Preventive War and American Democracy.* New York: Routledge.

Slantchev, Branislav L. 2011. *Military Threats: The Costs of Coercion and the Price of Peace.* Cambridge: Cambridge University Press.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43 (4): 1254–83.

Su, Chi. 2008. *Taiwan's Relations with Mainland China: A Tail Wagging Two Dogs.* London: Routledge.

Tellez, Juan, and Jordan Roberts. 2019. "The Rise of the Islamic State and Changing Patterns of Cooperation in the Middle East." *International Interactions* 45 (3): 560–75.

Van Evera, Stephen. 1999. *Causes of War.* Ithaca, NY: Cornell University Press.

Waltz, Kenneth N. 1979. *Theory of International Politics.* Reading, MA: Addison-Wesley Publishing Co.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47 (4): 363–75.

Weidmann, Nils B., Doreen Kuse, and Kristian Skrede Gleditsch. 2010. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36 (1): 86–106.

Weller, Nicholas, and Jeb Barnes. 2014. *Finding Pathways: Mixed-Method Research for Studying Causal Mechanisms.* New York: Cambridge University Press.

Williford, George W., and Douglas B. Atkinson. 2020. "A Bayesian Forecasting Model of International Conflict." *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 17 (3): 235–42.

World Bank. 2020. "World Development Indicators." Accessed July 4, 2021. https://databank.worldbank.org/source/world-development-indicators.

Wu, Sofia. 1995. "Chief Says PRC Missiles Able to Reach Island." Taipei CAN in English, February 24, 1995 in FBIS-CHI-95-037, February 24.

Yarhi-Milo, Keren. 2014. *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations.* Princeton, NJ: Princeton University Press.

Youden, William J. 1950. "Index for Rating Diagnostic Tests." *Cancer* 3 (1): 32–35.

Zegart, Amy, and Michael Morell. 2019. "Spies, Lies, and Algorithms: Why U.S. Intelligence Agencies Must Adapt or Fail." *Foreign Affairs* 98 (3): 85–97.

Zhu, Xuewen, and Lu Qibin. 1995. "Acquiring Masterly Skills through Intensive Exercises in Unfamiliar Sea Area: Certain Navy Unit Carries Out Combined Services Offensive and Defense Exercises, Using Various Kinds of Warships and Fulfilling Different Tasks." FBIS-CHI-95-065, April 5.

Zou, Kelly H., A. James O'Malley, and Laura Mauri. 2007. "Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models." *Circulation* 115 (5): 654–57.