What is Escalation?: Measuring Crisis Dynamics in International Relations with Human and Machine Generated Event Data

Word Count: 8923

# Abstract

When a dangerous international crisis begins, leaders need to know whether their next move is going to resolve the dispute or amplify it out of control. Theories of conflict have mainly served to deepen the confusion, revealing fighting, bargaining, and signaling to be high-dimensional and subtle equilibrium behaviors with deeply contextual consequences. Should a leader communicate resolve through aggressive acts, avoid spirals through accommodation, or focus on ensuring the possibility of a bargain? We offer a data-driven empirical solution to this logjam in the form of a new large-scale analysis of actions taken within 475 crises. We combine two complimentary measurement projects, the human-coded International Crisis Behavior Events (ICBe) dataset and the new machine-coded ICBeLLM. We model directly whether an action tends to shorten or extend the length of a crisis. The result is a directly interpretable measure of the latent escalatory/deescalatory nature of each action leaders have chosen over the last century.

# 1.0 Introduction

What does it mean that an international crisis is escalating, and what actions within a crisis are escalatory? This question haunts recent and ongoing crises such as missile tests on the Korean Peninsula, Russian incursions in Ukraine, partial state death in Syria, and Chinese claims of sovereignty in the South and East China Seas. In these situations, leaders face a shared fundamental question about how to navigate a dangerous situation without accidentally sliding over the line into a larger-than-desired conflict (Snyder and Diesing 1977; Slantchev 2005). Often leaders must make decisions with limited time and inadequate information (Jervis 1976, p. 95). What guidance can we offer to help maintain peace and stability? Can we predict how states will react to specific events during crises? Can we recommend policies or actions that are most likely to lead to success without being destabilizing? In short, what empirical evidence can we bring to bear on the practical questions of crisis management?[[1]](#footnote-21)

The international relations field offers two foundational perspectives on crisis dynamics, classically described by Robert Jervis (1976) as the deterrence model and the spiral model. The former blames escalation on the failure to credibly communicate threats (Rovner 2020) or misperceptions about capability or resolve (Morrow 2019). The latter views escalation as a retaliation to provocation where a tit-for-tat response fails to end the crisis (O’Neill 1991). These models offer contradictory policy prescriptions: on one hand, policymakers should use bold words and actions to discourage aggression; on the other, they should mitigate security dilemmas through compromise and accommodation. Historical studies typically provide more nuance, but the additional context needed to make sense of specific crises just highlights the fact that both models are underspecified. Neither model, moreover, is well integrated with more recent thinking about the underlying bargaining dynamics of military contests (Fearon 1995). Previous empirical research designed to validate or reject competing theoretical explanations for escalation has faced daunting challenges given that the basis for comparison and/or the logical discrepancies between alternative theories have not been resolved (Glaser 1992; Kydd 1997; Zagare and Kilgour 1998). These efforts have yet to establish a consensus pecking order between canonical alternatives (c.f., Wright 1965; Carlson 1995; Kinsella and Russett 2002).

Recognizing the enormous efforts of previous theoretical and empirical scholarship, we adopt an inductive approach designed to explore the correlates of escalation/de-escalation in a novel data source constructed by the authors and others (Douglass, et al. 2022). Our approach capitalizes on a new dataset, the International Crisis Behavior Events (ICBe) dataset, augmented by machine-coded data ICBeLLM. To our knowledge, these are the first international conflict data to code the full sequence of statements and actions by actors necessary to capture escalation as a process. These data allow us to track the actions and reactions of every relevant actor in the 475 interstate crises available from the canonical ICB dataset; who did what to whom, when, where, and how. Doing so allows us to identify various factors most likely to lead to escalation or de-escalation in crises. Using these triggers, we can develop a model that estimates (predicts) both the path and duration of international crises and that indicates likely precursors and their effects. We find support for neither the deterrence nor spiral model, in their classic forms, suggesting the need for a more nuanced theory that can accommodate the historical contingency and complexity of crisis processes.

This analysis is organized as follows. We first briefly review the most relevant examples of research on deterrence and the spiral model, both theory and evidence. We then discuss the problem of inference about processes utilizing non-process datasets (the only data that has been available to researchers up until this point in studying international conflict). The bulk of the chapter focuses on our inductive empirical “deep dive,” presenting details of the novel ICBeLLM dataset (Douglass, et al. 2023), identifying variables associated with escalation and de-escalation, and estimating our crisis duration model, based on these variables. We conclude with some insights of the project and suggestions for future research.

# 2.0 Crises, Escalation, and the State of the Art

The scientific state of the art on crisis escalation is too large and amorphous of a topic to fully summarize here, and so we draw attention to the minimal necessary machinery for understanding the empirical argument made here. Those components are the domain of the activity in question, the definition of the outcome we seek to explain, theoretical priors about the connection between different possible actions and that outcome, and empirical attempts to measure the above.

## 2.1 Crisis and Stability

The status quo in international relations is almost always one of stability, slow-changing bargains, and relative cooperation. War and conflict are the exceptions (Vasquez and Valeriano 2010). Formally, consider a set of autonomous political actors or players, 𝑝𝑃, consisting of states, international organizations, and subnational organizations. At each moment in time, , they engage in behaviors, $b\inB$, individually and towards each other which they each then hold preferences over. Any player can transition to a period of crisis when those behaviors swing toward “disruptive interaction with a heightened probability of military hostilities that destabilizes states’ relationships or challenges the structure of the international system” (Brecher and Wilkenfeld 1982). The International Crisis Behavior project has documented 496 such periods since 1918 (Brecher and Wilkenfeld 1997). Formally, we can think of this as a simple Markov model with two discrete latent states, stability and crisis , . This latent state is in turn partially observed in the current and historical behavior of the players. The measurement strategy for the unobserved latent state is to use observables like behaviors as proxies, $b\_{p^-1,t}=F(\_{p,t}) $. When other players are behaving badly, we think a player is strictly more likely to be in a state of crisis. However, the policy-relevant question we want to answer concerns the causal impact of a behavior on the continuation or resolution of a crisis, . This reveals an immediate conceptual difficulty- we are using measurements of behaviors observed in time t to predict a phenomenon in . The definition of and measurement of a discrete-time step within a crisis is doing the majority of our conceptual work. This fact has created unending confusion between measurement, explanation, prediction, and prescription in the study of crises, and we tackle it directly below.

## 2.2 Defining Crisis

For now, imagine a stylized crisis with two discrete and clearly distinguishable time steps, t and t−1. In t, we want to know whether a player is in a state of stability or crisis. We observe the behaviors of other actors in time t, 𝑏𝑝−1,𝑡. International relations theory holds strong beliefs about which behaviors are more crisis-like. For each behavior, b, assign a weight w=w1,w2,,wn, representing the strength of the signal it provides that a player is currently in a crisis state and the discrete state is some function of the sum of each of these weighted observed behaviors p,t=F∑i=1nwibi.  
We draw prior beliefs from international relations literature about the sign and rank ordering of these weights across behaviors. Much of 20th-century security studies was preoccupied with conventional warfare as the most extreme behavior, e.g. the Correlates of War (COW) project. In the shadow of possible nuclear war, scholars turned to measuring events that might be a precursor to actual fighting, e.g. the threat, display, or use of military force short of war recorded by the Militarized Interstate Disputes (MIDs) dataset (Maoz, et al. 2019). Others recorded entire streams of events from news headlines (McClelland 1978; Goldstein 1992; Azar 1993; Sherman 1994) and defined their level of departure from stability by hand, e.g. Goldstein scores (Goldstein 1992), or through unsupervised dimensionality reduction, e.g. Item Response Scores (Schrodt 2007). The ICB project took an even wider view, recording a “trigger” for each crisis which could take one of 9 different forms (verbal act, political act, economic act, external change, other non-violent act, internal verbal or physical challenge to regime or elite, non-violent military act, indirect violent act, or violent act).   
The ICBe ontology divided crisis events into two conceptual categories across armed actors vs unarmed actors and, given that, escalatory/uncooperative vs de-escalatory/cooperative behavior, shown in the table below. It further recorded details related to scale and severity, like the number of troops involved in an act, number of casualties, amount of territory exchanged, etc. The labels of escalatory/de-escalatory were perhaps premature, and instead for our purposes here we argue these are better thought of as indicators for and against a state of crisis/stability within a single discrete time step.

1. Consistent findings in the context of cross-domain deterrence and escalation using the ICBe data are described in Gannon (2022). [↑](#footnote-ref-21)