

Leadership importance, institutional constraints and conflict: A
document-based approach.

June 17, 2022

Abstract

At what point during conflict are individual leaders important decision-makers? Scholars of leadership, domestic institutions and conflict have debated this question for decades, but an answer is elusive because we lack a systematic measure of leadership importance. Exploiting recently declassified CIA President's Daily Briefs we construct a cross-national, weekly measure of how foreign policy experts perceive leadership importance that weights a leader's contribution to the policy-making process relative to other domestic actors. We reach two findings after estimating the determinants of leadership importance in over 16,000 statistical models that balance different inferential concerns. First, leaders are not perceived to matter equally over every stage of conflict. They matter the most during crisis negotiations when conflicts can either de-escalate to peace or escalate to war. They matter some during armed conflict and at conflict termination. They are not perceived to matter any more during conflict instigation than they are during peacetime. Second, domestic institutions designed to constrain leadership choices influence a leader's perceived role in the policy-making process. Leaders of heavily constrained regimes are perceived as less important day-to-day than leaders of unconstrained regimes. Further, leaders of constrained regimes are no more important at any stage of the conflict process than they are in peacetime. Thus, findings that suggest that foreign policy experts believe leaders matter a lot during conflict on average are based on the localized effects of unconstrained leaders during certain phases of conflict.

Word Count: 11,669

Since [Waltz \(2001\)](#) wrote *Man, the State, and War*, researchers have fiercely debated: How important are leaders for explaining international conflict? Recently, researchers have collected cross-national data on leadership tenure ([Goemans, Gleditsch, and Chiozza, 2009](#)) and pre-tenure characteristics ([Ellis, Horowitz, and Stam, 2015](#); [Fuhrmann, 2020](#); [Gift and Krmaric, 2017](#)), as well as leadership actions ([Lindsey and Hobbs, 2015](#)). Using this data, scholars argue that leader-level variation generates uncertainty about how states will behave in a crisis. The disconnect between how a government perceives a foreign leader and how that leader behaves creates opportunities for conflict ([Yarhi-Milo, 2014](#); [McManus, 2019](#)). In the aggregate, these quantitative studies provide convincing evidence that leaders exercise discretion in crises, and uncertainty about that discretion influences conflict *on average* ([Horowitz and Fuhrmann, 2018](#); [Chiozza and Goemans, 2011](#)). It is no longer reasonable to say that structural factors, or domestic institutions are so powerful that leaders cannot systematically influence conflict (For debate, see [Milner and Tingley, 2015](#); [Polsky, 2013](#); [Saunders, 2011](#); [Downs and Roche, 1994](#); [Mearsheimer, 2001](#)).

Even if leaders matter for conflict outcomes *on average*, it does not mean leaders are important during every stage of conflict. Although not always described this way, studies that find leaders matter theorize that leaders are only important at a specific stages of the conflict process. Some argue that leaders raise the chance a crisis will start ([Bak and Palmer, 2010](#)), others focus on escalation from a crisis to violent conflict ([Gottfried and Trager, 2016](#)), and others still argue that leaders influence how armed conflict is carried out or ends ([Krmaric, 2018](#); [Min, 2021](#); [Goemans, 2000](#)). While all studies linking variation in leader traits to conflict outcomes are plausible, they may not all be true ([Krmaric, Nelson, and Roberts, 2020](#)). It is possible that domestic institutions and structural factors constrain leaders considerably during specific stages of conflict. If true, then quantitative studies may find that leaders matter for conflict *on average* even if leaders are irrelevant for the particular choice that theory is focused on.

We shift the literature in a new direction by considering not which leader traits matter, but *when* they matter. In order for leaders' individual traits to express themselves in conflict outcomes leaders must have policy discretion. Therefore, we want to know: (1) at what stages during the conflict process do leaders make—or are perceived to make—important choices? And (2) do institutions uniformly constrain leaders at every decision-node, or do they constrain more at certain stages of conflict? These questions change the focus from *How* constrained are leaders to *What* choices are

constrained. We are not the first to recognize that these are critical unanswered questions (Jervis, 2013; Fuhrmann, 2020; Chiozza and Goemans, 2011). But, as others point out, the available data does not lend itself to answering them (Krcmaric et al., 2020).

We gain insight into these questions through a novel, time-varying measure of perceptions of foreign leader importance for foreign policy decision-making. Our measure exploits 4993 recently declassified, daily reports from the CIA to the president of the United States: the President’s Daily Briefs (PDBs). According to elite interviews with producers (CIA staff) and consumers (NSC policymakers) of intelligence products, and our review of the PDB’s historiography, the PDB functions as a top secret newspaper for the busiest person in the world (cf Priess, 2016; George and Bruce, 2008; Davis, 1995). By design, it is concise, written in simple English, and only reports enough information to explain what events just occurred around the world. The PDB is an ideal source to measure the CIA’s perceptions of leadership importance: when the CIA perceives that foreign leaders do not significantly contribute to state-actions, the CIA reports on the event but omits information about the leader. Using machine coding and human validation we construct a weekly measure of US perceptions of leader importance based on the number of times the CIA discusses a foreign leader by name relative to the number of times they discuss a foreign country or capital city by name.

Next, we conduct an extreme bounds analysis that estimates over 16,000 statistical models. The models treat phases of conflict, leader attributes, and institutional variables as predictors of perceived leadership importance. We draw inferences by contrasting the perceived importance of foreign leaders during peace-time and at distinct stages of the conflict process.

We establish our measure’s construct validity taking into account well-documented cases where cognitive and organizational biases have influenced high-profile intelligence reports (cf Jervis, 2010). We argue that the PDB’s focus on current intelligence (Kovar, 2000), the authors’ incentive structure (Robarge, 2005), the extensive database the CIA maintains on potential future foreign leaders (Dyson and Duelfer, 2020), and the PDB’s production process (Marrin, 2012) all make the PDB less susceptible to biases on leadership questions than strategic intelligence products (Blight and Welch, 1998). We also point out that CIA assessments are accurate on average despite a few high-profile cases of bias (Mandel and Barnes, 2014, 2018). Nevertheless, our inferences are not based on our measure alone. Rather, we exploit temporal differencing in our statistical techniques

to net out systematic biases. We explain how our overall approach addresses broad categories of potential confounds by walking through several examples: presidential pandering, writing style, redacted text, psychological biases, and variation in the frequency of events (e.g. speeches). We also illustrate our statistical findings, and the accuracy of our measure, through a case on the 1967 Six Day War.

Of course, no empirical approach is perfect. However, ours has two advantages over models that estimate conflict outcomes directly. First, conflict models face selection concerns (Krcmaric et al., 2020). The most salient is that each step in leadership selection and conflict escalation is part of an interactive strategic process (Wu and Wolford, 2018). We do not face these issues because the choices in foreign countries are not influenced by daily CIA reports to the president. Second, past scholars search for how a specific characteristic (e.g., education) influences a specific conflict outcome (e.g., crisis escalation or not) (Horowitz and Fuhrmann, 2018). But they struggle to disentangle the effects of specific biographical features and institutional constraints on the phases of conflict. We treat the timing of each conflict phase, leader attributes and regime features as predictors of importance in each model.

By side-stepping these inferential challenges, we bring evidence to bear for one question at the center of the leadership-structure debate: on average across historical cases, at what point of the conflict process are leaders important decision-makers? We find that leaders are perceived as the most important when states must choose between escalating a crisis to violent conflict or deescalating to a negotiated settlement. Leaders are seen as only moderately important during ongoing combat and during settlement negotiations that end combat. In contrast, leaders are not perceived to be more important for initiating crises than they are during peacetime. We observe similar patterns during analogous stages of civil conflict, giving us additional confidence in our findings and suggesting an exciting avenue for future research. These results will help scholars of diplomacy (cf Trager, 2010), and leadership biography and tenure (cf Krcmaric, 2018; Fuhrmann, 2020; Wolford, 2007) focus their effort on the specific decision-nodes where leaders are perceived to matter the most. It seems that structural factors like incentives (Mearsheimer, 2001) or accidents (Schelling, 1957) explain why crises start, but leader-specific variation explains why and whether they escalate to conflict or not.

We also address debates about the domestic constraints faced by leaders at the heart of studies

of domestic institutions and conflict. At the extremes, our findings are consistent with institutional accounts. Leaders of regimes with the strongest constraints are perceived to be no more important during any phase of conflict than they are during peacetime. In contrast, leaders of regimes with the weakest constraints are more important at almost every stage of the conflict process. Between these extremes, marginal effects show that leaders are important for crisis escalation in moderately constrained regimes. However, even moderate constraints render leaders not significantly important for combat operations and combat termination. This fits with accounts that leaders have freedom of action during high-stakes diplomatic negotiations (Yarhi-Milo, 2014), but are constrained when making choices that require expertise, including the conduct of military operations (Grynaviski, 2018). It also fits with political economy research that seeks to disentangle constraints from regime type (Henisz, 2000; Bütthe and Milner, 2008; Boix, Miller, and Rosato, 2013; Graham, Miller, and Strom, 2017) and also research that finds specific domestic institutions constrain specific choices (Mansfield, Milner, and Rosendorff, 2002; Milner and Tingley, 2015). We detail the implications of these insight for conflict.

Beyond leaders, domestic institutions and conflict, we clarify differences between current and strategic intelligence products (Priess, 2016; Friedman, 2019). While these differences are well known in the intelligence literature (Marrin, 2012; Blight and Welch, 1998; Lowenthal, 2019), we explain their implications for conflict driven by misperceptions. This distinction also helps us understand the conditions under which states can cultivate reputations (Kertzer, 2016; Schelling, 1957) by explaining when intelligence agencies can make strong inferences from a state's militarized choices. It also helps characterize different challenges and opportunities for scholars who quantify different declassified national security documents (Katagiri and Min, 2019; Lindsey and Hobbs, 2015; Lindsey, 2019; McManus, 2019). We discuss policy implications in the conclusion.

1 Are leaders important during conflict?

We define **leadership importance** as the extent to which a leader exercises discretion over state policy. As we explain more below, a leader is most important when her choices are unconstrained by other domestic actors (such as other elite, lobby groups, or the public). She is least important when the configuration of domestic institutions forces her to compromise extensively over the choice

she makes.¹

In our analysis, we distinguish between actual and perceived leadership importance. The former is the objective truth. The latter is the extent to which foreign policy experts believe leaders matter for making a specific kind of choice. To the extent that impressions of leaders influence strategic interactions, both perceptions and actual importance are critical to explain conflict outcomes (Krcmaric et al., 2020; Bak and Palmer, 2010; Koch and Fulton, 2011). We accept that some systematic bias may exist that cause perceptions and reality to deviate on questions of leadership (cf Yarhi-Milo, 2014). In sections 2.3 and 2.4 we discuss the inferences we can draw by balancing these concerns against our unique data and statistical method. For ease of exposition, we summarize the literature in terms of leadership importance generally.

Leadership importance in foreign policy outcomes may vary throughout a leader’s tenure. For example, leaders may matter a lot when they first come to office (Chiozza and Choi, 2003), when states start a nuclear program (Fuhrmann and Horowitz, 2015), or pursue certain economic policies (Besley, Montalvo, and Reynal-Querol, 2011).² Yet many scholars have focused on the role of leaders during conflict because they believe leaders are more important during conflict than they are on average throughout their tenure (Byman and Pollack, 2001; Milner and Tingley, 2015). There are two reasons. First, crises are fast-paced events. States do best in crises when they can quickly capitalize on opportunities. Recognizing that time is of the essence more in a crisis than during, for example, welfare reform, governments are willing to forego oversight. Second, all factions of government work together in the face of a foreign threat. To the extent that politics stops at the water’s edge, especially during a crisis that risks breaking out into war, different domestic factions will allow a leader to exercise discretion (Milner and Tingley, 2015). As a result, we focus on differences between leadership importance during each stage of conflict and peacetime.³

Despite these accounts, there is some debate about whether leaders are especially important during conflict. Leaders rely on advisers because they lack foreign policy expertise (Grynaviski,

¹To be clear, our definition and framework does not rely on the fact that different leaders would make different choices. Rather, it emphasizes differences in leader discretion relative to other domestic actors at different decision-nodes. In our empirical analysis, it does not matter what a counter-factual leader would have done. The CIA may focus on leaders more at times of high importance because the leader’s preferences matter more relative to other domestic actors. The fact that different domestic leaders could make different choices is complimentary to our theory because it creates additional uncertainty that makes learning about leaders at times of high importance more critical.

²We control for these events in our empirical analysis.

³Cross-time comparisons helps us address inferential challenges. Our results are identical if we compare importance across phases of conflict or interpret each phase on its own terms.

2018; Saunders, 2017; Dyson, 2009; Lindsey, 2017), and do not have complete control over how crises emerge and evolve (Schelling, 1957). Further, powerful structural incentives may render individual preferences redundant during these high-stakes events (Mearsheimer, 2001).

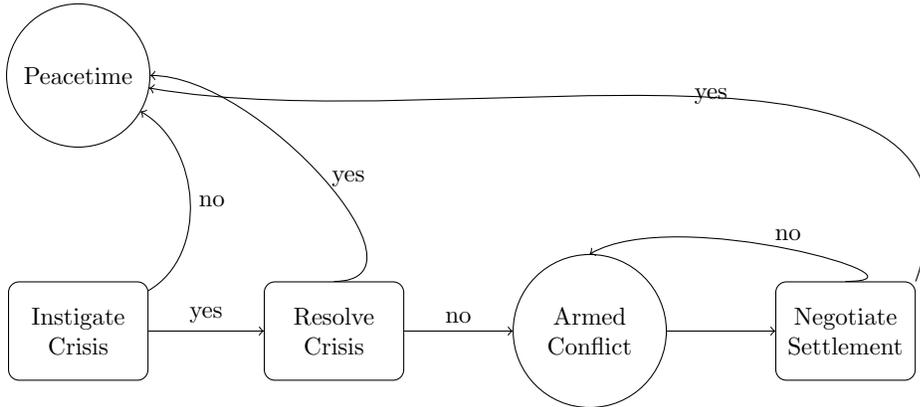
An enduring criticism against the importance of leaders relies on insights about domestic political institutions (Kaelberer, 1997; Mesquita, Smith, and Siverson, 2005; Jervis, 2013). There are two ways that institutions could render leaders less important during conflict. First, institutions prevent leaders from exercising discretion through veto points (Polsky, 2013). Second, institutions can punish leaders who exercise discretion poorly. The expectation of punishment can constrain a leader's choices (McGillivray and Smith, 2006; Croco and Weeks, 2016; Reiter and Stam, 2002).

1.1 Existing evidence and decision-nodes during conflict

Modern debate focuses on leader discretion, which leader attributes influence behavior, and which sorts of regimes can moderate leader importance. To address these questions, researchers have collected extensive data on leadership characteristics (Krcmaric et al., 2020; Goemans et al., 2009; Horowitz and Fuhrmann, 2018). They correlate variation in leadership characteristics with choices that leaders make during the conflict process to evaluate if leaders are important for explaining conflict outcomes. Researchers have identified four phases during the conflict process where leadership variation predicts an important aspect of conflict. We depict these four choice-nodes in Figure 1. Each box represents a decision node where leaders *might* exercise discretion. One line of scholarship argues that leaders directly instigate crises by making territorial demands, or provoking a foreign country. To test leaders' role in starting crises, researchers code the onset of a non-violent crisis (or the absence of crisis) as the outcome of interest (often as the onset of a non-fatal militarized dispute). A second research program holds that leaders exercise discretion in how they negotiate during an ongoing crisis, which affects how crises are resolved. Scholars code whether a crisis escalates into fatal conflict (the onset of a fatal MID), or de-escalates (the end of a non-fatal MID) leading back to peace as the outcome of interest.

A third group argues that leaders affect the conduct of ongoing armed conflict by choosing military strategy and force commitments (leaders matter in the days that a fatal MID is ongoing). They focus on states already involved in a fatal militarized dispute, then code whether those states win or lose battles. A fourth group argues that as the outcome of war becomes clear, leaders choose

Figure 1: The Conflict Process



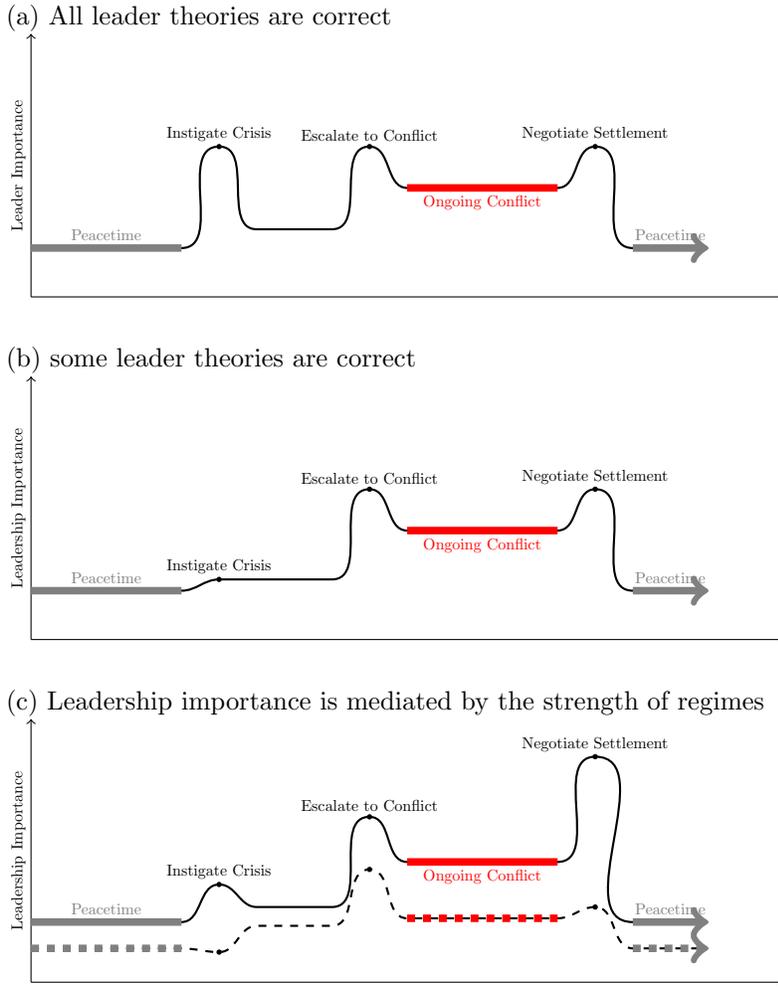
Blocks represent phases of the conflict process. Circles represent states of the world based on these choices.

whether and how states negotiate.

Using this empirical strategy, researchers have reported scores of significant findings relating different leadership characteristics to state choices at a given conflict-node. In the aggregate these studies are convincing. But individually, each suffers from two selection concerns (Krcmaric et al., 2020). First, conflict is a strategic process. The outcome of one stage influences what other stages we observe, and also how states act at these future stages. How states anticipate their own or their counterpart's performance at a future stage influences the choices they make today. Empirical models may observe a leadership characteristic with a significant coefficient at one stage of the conflict process not because leaders exert influence at that point, but because that leader will make important choices at a different point.

Second, personal characteristics are not randomly assigned. People with certain traits select into specific experiences. For example, people with a high risk-tolerance may choose a career in the military. Furthermore, institutions that select leaders anticipate the problems a leader will face and choose accordingly. Thus, leaders with specific characteristics may be more likely to face conflict. If true, specific characteristics may positively predict conflict outcomes only because leaders with those characteristics are selected when conflicts are likely to end that way. As a result, it is difficult to know exactly the conflict stage at which leaders matter.

Figure 2: How perceived leadership importance would vary across conflict stages if:



Each panel represents leadership importance at different stages of a conflict. The higher the curve at a certain point, the more important the leader is at that phase of conflict. In panel (c) the dashed line represents leadership importance in states with strong regimes. The solid line represents weak regimes. On average, the curves in panel (c) add up to panel (b).

1.2 Do leaders matter for certain choices more than others?

We argue that studies linking leader characteristics to specific phases in the conflict process have skipped an important step. Before linking leader characteristics to conflict stages, we need to know the stages at which leaders exercise, or are perceived to exercise, the most discretion.

We visually represent variation in leadership importance at different conflict decision-nodes in Figure 2. Each panel represents a single crisis that escalates all the way to conflict, then is resolved. The y-axis is how important the leader's personal choices are relative to other factors. Panel (a) represents the cumulative predictions of the leadership literature. According to these studies, leaders are more important at every stage of conflict than during peacetime.

Our first point is that the general criticisms of the leadership literature may not uniformly apply at every decision node in the conflict process; but still may apply forcefully at specific stages. Panel (b) represents one *possible* alternative to the aggregate findings of the leadership literature. In panel (b), general criticisms of the leadership literature apply unevenly across stages of conflict. Leaders may be subject to forces beyond their control more frequently at some points than others. For example, President George W. Bush played no part in starting the Hainan Island crisis, which began with a mid-air collision between an American and a Chinese aircraft. However, once the crisis had started, Bush had some control over whether the crisis ended peacefully or not. Related, leaders may vary in how much they rely on advisers at each decision point. They may rely on advisers to manage ongoing military operations during combat but unilaterally decide how to negotiate with foreign counterparts. Thus our first question is: *at what conflict stages are leaders important for making state decisions?*

Our second point is that institutions mediate leadership importance at specific stages of the conflict process. Thus, even if leaders can overcome institutional constraints on average across a conflict, other domestic actors may powerfully constrain leaders at a specific decision-node. Thus our second question is: *at what conflict stages do institutions mediate leaders' importance for making state decisions?*

We visualize the potential role of institutions in panel (c), which builds on panel (b). The example dis-aggregates regimes based on the extent of constraint. The solid line represents regimes with strong institutions that constrain and punish leaders. The dotted line represents institutions that struggle to constrain. Consider the point at which leaders must choose whether to terminate an ongoing conflict. On average across all regimes (panel (b)) we represented that leaders made important choices at this point. But it is possible that the effect is driven by weakly constraining regimes in which leaders matter a lot. If true, the conclusion that leaders matter a lot on average in spite of political constraints does not explain the data well.

Now consider the point where leaders select into crises. On average across all regimes (panel (b)) we represented that leaders did not make important choices at this point. But it is possible that leaders of weakly constraining regimes do matter more at these points than they do during peacetime while leaders of highly constraining regimes matter less. If true, then variation in institutional constraints can recover a finding we lose on average.

To be clear, these panels are just examples. We see good theoretical reasons why leaders would be very important (or not) at each stage of the conflict process, and also why constraints would mediate leadership importance at each stage. Our point is that leaders are unlikely to matter equally across regimes or stages of conflict. Until we figure out when leaders matter the most, we cannot link specific leadership traits to specific outcomes at specific stages of the conflict process or properly test signaling theories.

2 Empirical Approach

We use a novel data source, the CIA’s President’s Daily Brief, to construct a cross-national, temporally granular (weekly) proxy measure for perceived leadership importance. Treating this measure as the outcome, we then estimate how important a leader was perceived to be on weeks surrounding different phases of the conflict process while controlling for thousands of combinations of other events, regime and leader attributes and country fixed effects. We draw inferences by comparing parameter estimates that tell us the relative importance of leaders at each phase of conflict and at peacetime.

As discussed in the introduction, our unique data source affords us many advantages that allow us to side-step inferential concerns common in studies that estimate events (such as conflict onset). However, our data raise different concerns. Table 1 summarizes three categories of potential confounds that may concern readers and how we address them. Most concerns take one of two forms: our quantitative measure may not capture the intended meaning as it is written in the documents; the information in the reports themselves could be systematically biased.⁴ Our main defense is that our inferences are based on a statistical technique of *temporal difference* and not our measure alone. So long as our measure correlates with perceptions of leader importance on average across all crisis episodes, then CIA bias or bias brought on by how we convert CIA reports into our measure only threatens our inferences if it affects reporting *differently* at each stage of conflict and at peacetime. We walk through the logic of this defense in the context of cognitive biases and presidential pandering in section 2.4.

After a review of the international relations and intelligence literature, as well as interviews

⁴Other proxies based on expert assessments (e.g. Polity (Marshall, Jaggers, and Gurr, 2002)) share this issue.

Table 1: Threats to validity and how we address them.

Concern: Your measure may not capture true perceptions of leader importance for every country-week.	
For example	(a) The CIA may report filler. (b) The measure only uses leader, country and capital city names.
How we address it:	
(1) Our inferences are based on the average estimate of importance across every phase of every crisis; not the measure’s accuracy for every country week.	
(2) Our coding procedure was developed in consultation with former PDB authors to raise the probability it is correlated with their true intent.	
(3) Human coders read original PDBs and scored leadership importance on that day. Their answers are highly correlated with our measure.	
Concern: PDBs may be systematically biased away from true leader importance/CIA perceptions thereof.	
For example	(a) CIA may respond to presidential preferences/bureaucratic incentives. (b) Cognitive biases cause PDB authors to under/over-report on leaders. (c) Redacted text means there are paragraphs that you do not consider
How we address it:	
(1) We draw inferences by contrasting the parameter estimates of different periods of conflict and peacetime. Thus, biased reporting only confounds our inferences if it increases/decreases leader mentions at one stage of conflict but not others. Consistent, systematic under/over-reporting on leaders does not damage our inference.	
(2) The use of a ratio-based measure nets out spikes in discussion caused by conflict.	
Concern: Differential biases.	
For example	(a) A PDB author’s bias may be activated by conflicts between two foreign states. (b) Foreign leaders make more public comments at conflict onset. The PDB always reports what foreign leaders say.
How we address it:	
(1) We estimate 16,000 regressions that include different combinations of time varying predictors. If our measure was differentially biased at each stage, we would see fluctuations in our parameter estimates and standard errors across models. We do not.	
(2) Human coding is highly correlated with our measure for PDBs written during peace-time and conflict. We would see differences between peace-time and conflict if differential bias was present.	
(3) Features of the PDB’s production process, the focus on current (not strategic) intelligence, and institutional rewards targeted at accuracy suggest that the PDB may be less susceptible to bias than other intelligence products.	
(4) Arguments about differential behavior of foreign leaders are consistent with the logic of our measure. For example, leaders give more speeches when those speeches can matter. If the speeches matter, then reporting on them accurately reflects a leader’s importance.	

with analytical methodologists at the CIA and intelligence historians focused on the PDB, we could only come up with three potential sources of differential bias: (1) conflict could be correlated with redacted text, (2) foreign policy events that are always reported (such as public statements by leaders) are more common at certain stages of conflict, and (3) the PDB authors’ institutional incentives may vary. We address each below. To bolster our confidence that we had not missed a confound, we exploit additional data validation checks. We describe three conditions under which we would observe heterogeneity in our data if our results were differentially biased. We show that our results are stable under these conditions. We also provide a historical review of the PDB’s production processes and other information that increases our confidence in our inferences.

In section 4, we examine the Arab-Israeli Six Day War to illuminate our overall findings in an

important case. But we also confirm that the CIA’s assessments of leader importance in Israel, Egypt and Jordan (and our measure) are consistent with the historiography.

2.1 The PDB as source material.

We generate our proxy measure using a single CIA intelligence product called the President’s Daily Brief (PDB).⁵ The CIA recently declassified PDB reports between June 1961 and January 1977. Over these 16 years, the CIA wrote the PDB on average 5.6 days a week (4993 documents).

We focused on the PDB for two reasons. First, the PDB is explicitly designed to report on all salient, ongoing foreign policy events. Richard Lehmann, who designed the PDB in 1961, recounted President Kennedy’s instructions: “[what I] need is something that will have everything in it that is worth the President’s attention, everything that is worth his knowing in all these things [lengthier intelligence products] so I don’t have to fuss with them.” Kennedy wanted “a single publication, no sources barred, covering the whole ground, and written as much as possible in the President’s language rather than in officialese... [and concise enough to] fit in his breast pocket” (Kovar, 2000). Given this broad coverage, the PDB rarely provides a detailed analysis of any single event. Instead, it primarily reports paragraph-long descriptions that highlight only salient information to help the president understand what happened and why.

In the context of our theory, to the extent that the CIA believes that a foreign leader is important for understanding why an important event (possibly a phase of conflict) unfolds the way that it does, the CIA will report on that foreign leader’s actions. If the CIA believes foreign leaders do not significantly contribute to that event, the PDB reports on the event but omits information about the leader.

Table 2 provides examples of how the PDB describes a fatal militarized dispute between India and Pakistan. Entry (a) describes Indian Prime Minister Indira Gandhi drawing a red line, the “annihilation” of Bengalis by West Pakistani forces, in Delhi’s standoff with Islamabad but omits detailed military discussions. Entry (b) focuses on military operations. There is no leader-specific discussion. The PDB ignores leadership information in case (b) because leadership choices are over-shadowed by military events. This is exactly the point of our measure. Military actions and observable leadership choices were ongoing on the days of both report (a) and report (b). However,

⁵Early version are titled the President’s Intelligence Checklist. We describe these as PDBs also.

the CIA only reported on leadership choices on days they were important. On other days during the crisis, the PDB did not report on leadership nor military actions because, presumably, neither the military nor the leader did something that day that the CIA believed was important.

Table 2: Example Entries from PDB Description of 1971 India-Pakistan War

(a) Leader relevant entry	(b) Leader irrelevant entry
12/1/1971: Both Indian and Pakistani officials are expressing growing pessimism on the chances for peace. . . Prime Minister Gandhi told the upper house of parliament yesterday that the presence of West Pakistani troops in East Pakistan constitutes a threat to India’s security. She warned that India would not stand by while Bengalis in East Pakistan were “annihilated.”	12/9/1971: The defense of East Pakistan is crumbling as Indian troops advance on nearly all fronts. The town of Comilla was captured yesterday, and the Comilla military cantonment is under seige. When that falls, of the three major army bases in East Pakistan—Comilla, Jessore, and Dacca—only Dacca will remain in government hands.

A second reason we focus on the PDB is that it is the intelligence product least likely to suffer from systematic bias. Research that documents psychological biases in intelligence assessments focuses on *strategic* assessments—such as estimates of a rival’s motives (Yarhi-Milo, 2014). The PDB reports *current* intelligence. Current intelligence is less susceptible to bias than strategic intelligence because it is focused on easily defined events that have already happened. In contrast, strategic intelligence assessments focus on predicting the future. As Heuer (1999) observes in his seminal work on biased intelligence, biases are more likely to afflict assessments on more complex topics, where analysts must balanced more unknowns, and therefore make more assumptions.

It is difficult to evaluate whether strategic assessments are systematically accurate because it takes many years to figure out what actually happened (Blight and Welch, 1998). Since managers cannot easily identify systematic error, it is difficult to adjust the production process. In contrast, the CIA regularly reviews the PDB. If the PDB repeatedly makes the same mistake on the same topic, managers adjust analyst training (Mandel and Tetlock, 2018) or the production process (Priess, 2016).

The PDB’s production process also makes it the product least likely to suffer from political or bureaucratic bias. The PDB was produced at the Office of Current Intelligence (OCI). OCI staff were career intelligence analysts—not political appointees—on rotation from the Directorate of Intelligence (DI).⁶ These staffers were “tasked with writing analytic summaries and other brief

⁶The OCI is an unusually demanding rotation because it requires 2 am start-times for many analysts. Thus there was frequent turn-over.

products for policymakers (CIA, 2008).”⁷ Appendix B reports the OCI’s organization chart, which visualizes the production process for the PDB. Every day, junior staff on regional desks read through intelligence products published over the last 24-48 hours including CIA published analytical reports, diplomatic cables, other processed intelligence, and open source information. Junior staff then short-listed relevant information, debated what was most salient, and passed it to the Publications Board, which is internally known as the “Front Office.” The Publications Board reviewed the material and discussed it with the Intelligence Staff (another organizational sub-unit), the original authors of the intelligence product, and the relevant regional teams. The PDB that was ultimately sent to the president was a consensus document that comes from this deliberative process. Like all employees, OCI staff faced bureaucratic incentives. However, evidence suggests that the OCI received its highest praise when it provided timely, accurate intelligence; even that went against the president’s prior expectations and against the advice of the president’s advisers (Robarge, 2005).

We focus on a CIA product because available evidence suggests that CIA assessments are accurate *on average* even if there have been high-profile instances of biased analysis. Specifically, Mandel and Barnes (2014) quantify the accuracy and confidence of over 1500 strategic assessments and find that “the results show that both discrimination and calibration of forecasts was very good.” If anything, CIA reports are less confident than they ought to be in the accuracy of their assessments (Mandel and Barnes, 2018), which may explain an over-estimation of the effects of systematic bias.⁸

Of course, this does not mean that the PDB is free from bias. As we discuss below, we address residual concerns through our statistical and inferential methods. We also validate our approach with several data tests.

2.2 Converting PDB text into a measure

We report details from our coding procedure in Appendix A. Following Katagiri and Min (2019); McManus (2019), and others, we convert text into a measure using a weighted count strategy. We separately count the number of times the PDB mentions a foreign leader, capital city, and country. We believe these three actors cover the level of analysis problem we want to measure. For example,

⁷This office was distinct from strategic intelligence, and operations (See George and Bruce, 2008, Chapter 1).

⁸The focus on cases of failure may also explain the over-estimation of bias (Marrin, 2012).

whenever a country chooses to fight, the CIA could say “Russia has decided...”, or “Khrushchev has decided...” depending on who they believe was most relevant actor. Informal discussions with former PDB briefers, and CIA archivists suggested we should also include capital cities. The reason is that the CIA often refers to broad political choices as coming out of a foreign capital (e.g., “Moscow has decided...”). We explain why we omitted other political actors and pro-nouns in Appendix A.5. We define countries and capital cities using the Correlates of War dataset, and identify leaders using ARCHIGOS 3.5.

Despite its name, the president only received the PDB 5 or 6 times per week. This made daily-level analysis difficult because there were NAs on days that conflicts started. To address this concern, we aggregated our counts to the country-week level (814 total weeks). We then constructed the following country-week outcome measure:

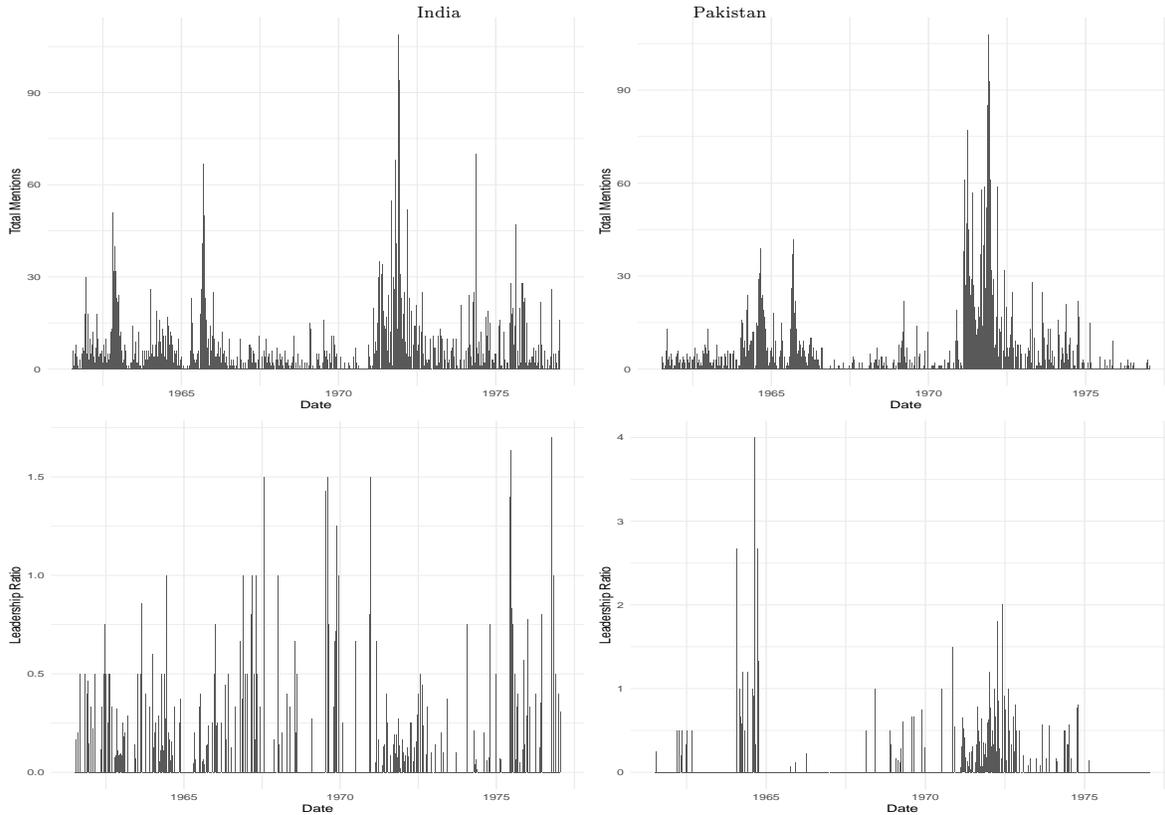
$$Leader\ ratio_{wc} = \frac{Leader\ mentions_{wc}}{country\ mentions_{wc} + city\ mentions_{wc}} \quad (1)$$

The subscripts indicate that we counted the total number of mentions of a leader, city and country (that share a given COW ID code) in a week. This measure increases when the PDB refers to the leader more relative to the country and capital. Because the measure uses a ratio, it accounts for country salience that may vary across time. For example, the PDB is likely to report a lot about a country during a crisis. But this will not increase our measure of leadership importance unless the PDB provides a lot of leader-specific information *relative* to discussion about the country’s behavior in general.

Figure 3 plots PDB discussion of India (column 1) and Pakistan (column 2) to illustrate temporal variation in the number of country mentions (row 1) and the leadership ratio (row 2) over our period. As the plots show, reporting is not evenly distributed. Rather, there are clusters of reports that correlate with important events. There is broad discussion of both countries across the entire period, generating several opportunities to mention foreign leaders.

However, the largest spike in country reporting in both India and Pakistan occurs during the Indo-Pakistan War (1971). The spike lasts longer in Pakistan, which may reflect their relations with Bangladesh over the period. There are also smaller shared spikes that both countries share that correspond with low-level MIDs in 1963 and 1964. But not all of the patterns are perfectly

Figure 3: Examples: India and Pakistan



The top row is total mentions of country, city and leader. The bottom row is our leadership ratio.

correlated across countries. There is more regular reporting on India (this corresponds partially with elections and leadership changes—which we control for), and also large spikes in 1963 we do not observe in Pakistan. In contrast, we see reporting on Pakistan in the late 1970s that does not correlate with reporting in India. The plots also show that country mentions and leader mentions are not perfectly correlated, suggesting variation in when the CIA believes it is important to report on leaders specifically as opposed to a country in general.

2.3 The leadership ratio closely approximates perceptions of importance

We argue that the leader ratio closely proxies the CIA’s perception of a foreign leader’s importance. To verify that our count-based strategy converted the CIA’s intended meaning into the leadership ratio accurately, we conducted a double-blind validation. We randomly selected (a) 25 PDBs written during an ongoing militarized crisis between two belligerents; and (b) 25 PDBs written on days where no MID took place. In case (a) we selected the two belligerents. In case (b) we

selected two countries mentioned in the document at random. We then asked an undergraduate student at [removed for review] to read the original PDF documents on the CIA website and answer the following question: *The document describes some beliefs, events or activities that relate to [Country]. At the time the document was written, the leader of this country was [Leader]. Based on the discussion in the PDB how important was the leader in the ongoing events, activities or beliefs that the CIA is describing.*

The student scored each entry on a five-point scale ranging from not important to important. We then computed the daily leadership ratio using the data our machine coding procedure produced. Our machine coded ratio correlated with the student coding 0.81 and 0.77 for case (a) and (b), respectively. The high correlation gives us confidence that our computed measure reflects what the CIA intended. The similar levels of correlation across conflict and non-conflict phases gives us additional confidence that differential bias is not inflating the ratio during conflict phases.

We now address three lingering concerns that the PDB does not reflect the CIA's best assessment, and may even differentially affect the accuracy of our measure at different stages of conflict. First, the CIA does not equally gather intelligence on all leaders and countries. As a result, they may want to report on a foreign leader but have information gaps. The missing data problem is amplified because analysts are subject to additional biases given the absence of information (Yarhi-Milo, 2014). Fortunately, during the period examined, the CIA extensively monitored potential future leaders for all countries and maintained a comprehensive biographic register (Dyson and Duelfer, 2020). Related, the CIA focuses more on certain countries. Fortunately, the PDB reports at least once on 141 of 149 countries; and 392 of 413 leaders who take office during the period.⁹

Second, the PDBs contain redactions that could influence our measure if they systematically exclude discussions about leaders more frequently. Fortunately, when the CIA redacts text, they provide a code that explains the reason for redaction. Over 90% of reactions in the document are for code 50X1 or 25X1. These codes represent that continued classification is legally justified by a need to protect sources and methods (i.e. who provided the information). Source redaction are not likely to contain discussions of the leader or the country and therefore will not bias our finding. Most other redactions cover complete entries (i.e. the entire report on a country for a given day). These

⁹Our statistical models include country fixed effects and an outcome variable that factors in both leader and country importance to address country-level heterogeneity.

complete-entry redactions leave the title of the entry (which is the country-name). One might worry that this inflates the number of country mentions relative to leader-mentions. However, we think this is unlikely to damage our inference for two reasons. One, complete-entry redactions occur only about twice a week. Two, complete-entry redactions, or redactions more broadly, are uncorrelated with the timing of conflict. Since they are not correlated with conflict, they are unlikely to create the differential bias necessary to damage our inference.

Finally, it is possible that foreign leaders are more active at certain phases of conflict. Since they are more active, the PDB authors have more leader events to describe. One concern is that leaders may engage in public diplomacy more frequently at conflict onset and termination and the CIA reports this. This concern is actually consistent with our theory. We know that public statement, visits, and even private diplomacy have important effects (Lindsey and Hobbs, 2015; Fearon, 1994). At times when leaders use these foreign policy tools more, they are more important. Of course, not all public diplomacy exerts influence. If the PDB reported every foreign policy statement then it could artificially inflate a leader's importance at times when speeches are common. We reviewed PDB and media reporting during the Indo-Pakistani War of 1971 and the Arab-Israeli Six Day War. We confirmed that the PDB often omits public statements reported in Western media outlets.¹⁰

2.4 Our inferences are robust to systematic bias in our measure: cognitive biases and presidential pandering.

Above we argued that the PDB is the least likely intelligence product to suffer from systematic bias. Of course, cognitive biases or perverse bureaucratic incentives may cause systematic error in our measure. We do not have space to review every potential source of bias (see Heuer, 1999, for review of some). Here we explain how our inferential method of temporal difference provides a common defense, and use two examples of bias to illustrate how our defense works.

We draw inferences from a statistical model that estimates differences in the leadership ratio over time. Our inferences do not assume that the CIA is unbiased in how it reports leadership

¹⁰E.g., On December 9, 1971, the *New York Times* reported that Indian Prime Minister Indira Gandhi rejected calls for a cease-fire in India's war with Pakistan, yet she was not discussed in the PDB. See <https://timesmachine.nytimes.com/timesmachine/1971/12/10/90705641.html?pageNumber=1>. And on June 7, 1967, the *New York Times*' James Reston reported Israeli Prime Minister Levi Eshkol's comments on Israel's war aims. The PDB does not mention Eshkol. On June 7, *Times* also reported diplomatic communications by both President Nasser and King Hussein. The PDB does not mention either of them on June 7. See <https://timesmachine.nytimes.com/timesmachine/1967/06/07/issue.html>.

information. Rather, even if a bias caused PDB authors to consistently over (or under) report on leaders, our inferences remain unbiased.

For example, [Yarhi-Milo \(2014\)](#) argues that intelligence analysts are less exposed to foreign leaders, and therefore do not rely on interactions with them in their assessments. She shows that intelligence analysts are likely to over-prioritize events they can observe directly. Thus, they under-report on leadership events because they cannot easily observe them. Suppose the CIA under-reports on foreign leaders because it has less exposure to them. This drives the leader ratio down during every stage of a crisis and peace-time. But our inference relies on *differences* in the leader ratio between these different periods. Thus, our inferences about true leadership importance are not impacted. To damage our inferences, a bias would need to reduce (or raise) the CIA's focus on leaders during a specific period of the conflict process and not other periods, or during peacetime.

As a second example, consider that PDB authors may have distorted information to curry favor with the president or respond to presidential requests. It may be the case that certain presidents want to hear more about leaders and the CIA responds. This will inflate our leadership ratio. However, the presidents' tenure is not correlated with the instances and timing of conflicts between two third party states. Thus, the president's preferences would raise the leadership ratio at every stage of conflict and peacetime. To ease any residual concerns about presidential request biases, we exploit president-level fixed effects. [Priess \(2016\)](#) notes that Nixon and Kennedy were at opposite ends of active listening: Kennedy interacted considerably with the PDB staff whereas Nixon never made requests. If presidential request substantially biased the documents, we would observe a statistically different result from a fixed effect for Kennedy and Nixon. We introduce presidential fixed effects and observe no statistical difference. Generally speaking, the fact that our results are robust to presidential fixed effects suggests that our results are robust to political biases and speaks to the overall robustness of our approach.

2.5 Independent variables: Conflict stages

Others have coded the different ways that stages of conflict can end (e.g. a crises can escalate to war, or de-escalate to peace). At root, these studies assume that there are critical decision nodes during the conflict process where leaders are important. Thus, our goal is to identify the weeks

Table 3: Converting choice outcomes into periods where leaders matter

State must choose:	Our importance measure based on:
Instigate Crisis or not.	Onset of non-fatal MID
Escalate Crisis to violent conflict or resolve it peacefully	Termination of non-fatal MID Point MID becomes fatal
How to command forces during conflict	Ongoing fatal MID
Whether/how to negotiate a peace settlement, or continue to fight	Termination of fatal MID

that are close to these key decision nodes. Our basic insight is that if a leader is important during this period, then the CIA should discuss the leader more at that point in time than they would at other points in time. We mainly rely on the MIDs dataset (Palmer, D’Orazio, Kenwick, and McManus, 2020) because MIDs are often the focus of the leadership literature, and because the coders include precise start and end dates for crises and armed conflict.

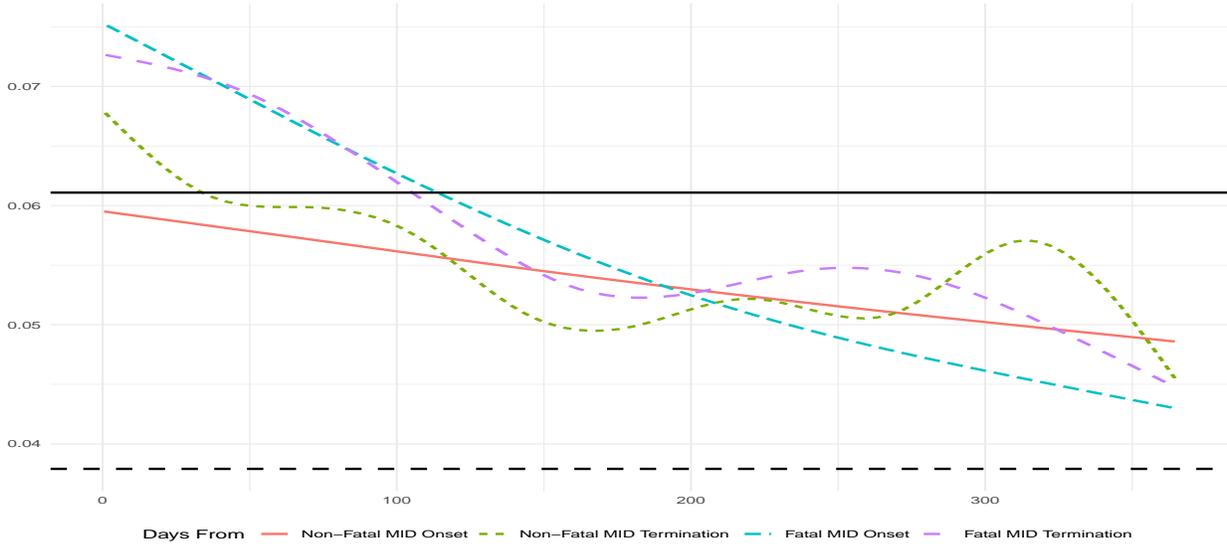
Table 3 summarizes the event dates we use to tie MIDs data to stages of the conflict process where leaders may be important. For every week, we code the time until the closest decision node for the onset and termination of non-fatal MIDs, the point that MIDs escalate to fatal MIDs, and the termination of fatal MIDs. For example, for the onset of a non-fatal MID, we take the start-dates of every non-fatal MID that country c is involved in and then measure:

$$\text{until nonfatal MID onset}_{cw} = \text{abs}(\min[\text{week startdate} - \text{MID startdate}]) \quad (2)$$

This measure is 0 if a country starts a non-fatal MID in the current week. Then as time moves away from the event start point (in either direction), the measure increases, until it starts to get closer to the onset of another non-fatal MID. The functional form ensures exponential decay. When correlated with our measure of leadership importance it weights the few days surrounding an event’s onset more than days that are further away.

Figure 4 plots lowest lines for our leadership ratio as a function of time until our four main potentially relevant events. Each line represents the leadership ratio score for a different event. The x-axis is the number of days from the events. At the point above day 0, the line summarizes the mean leadership ratio on the day that a critical event is happening. The black solid horizontal line is the mean leadership ratio score across all country-weeks. The black dashed horizontal line is the mean leadership ratio score in country weeks that a MID is ongoing.

Figure 4: Leadership Discussion Near Important Stages of Conflict



The mean leadership ratio for countries experiencing a conflict in a given week is 65% larger than countries not experiencing a conflict. Furthermore, all of the lines are downward sloping as they move away from 0. Consistent with existing accounts, this suggests that the CIA thinks leadership choices are more important during conflict than peacetime.

In the statistical analysis that follows, we construct the following measures for all of our event onset and termination dates. Using non-fatal MID onset as an example:

$$\frac{1}{\text{until non-fatal MID onset}_{cw} + 1} \quad (3)$$

If a non-fatal MID starts in the current week, the measure records a 1. The measure approaches 0 as the closest MID-week moves further away from the current week. We take the inverse for two reasons. First, the measure has an exponential-like feature so that the marginal rate of decrease is larger in weeks near the onset of a non-fatal MID. However, the measure does not change too much if a MID onset is months away. This matches our substantive knowledge of the PDB: the CIA discusses events as they happen. This inverse measure prioritizes discussion in the periods close to event onset. Second, since temporally distant MID approach 0, we can include countries that have never experienced a MID by coding every country-week as 0.

This measure does not require arbitrarily constructed time windows around events. It also

addresses measurement error in recording transitions between stages of conflict (Gibler, Miller, and Little, 2016). Still, one might worry that different stages of conflict are close. Even with an exponential time weighting, we may capture a relationship between our measure and crisis onset, for example, when leadership discussion is referring to crisis termination. To address this issue we estimate models that include all phases of conflict, each phase separately, then every combination of conflict stages. We find consistent results in every specification.

For countries that have a fatal MID ongoing in a week, we code an ordinal variable that adds 1 for each fatal MID ongoing in that country in a week. For example, if a country is fighting two wars simultaneously, we give it a score of 2 for that week. We also include shadow variables based on an alternative conflict process that should have similar effects. Specifically, the existing leadership literature has focused on leadership importance during international conflict. However, we see no reason that the mechanisms should not apply during civil conflict. Like international conflict, states must make choices about how to fight against and when to negotiate with sub-state belligerents. If a state's leader exercises discretion at these critical junctures in international conflict, they should also exercise discretion in civil conflict.

To wit, we measure time until the onset and termination of a fatal civil war, and days that a civil war is ongoing. Our hope is that when coefficients from the international conflict measures are significant, the corresponding civil war measures will also be significant in the same direction.¹¹ If this is what we find, the shadow measures will give us confidence that MIDs capture the different roles that leaders play at different phases of conflict, and that our results are not driven by some other feature of conflict nor by peculiarities in the MID coding scheme.

2.6 Independent variables: Institutional constraints

We measure institutional constraints on a leader's discretion using Henisz (2000)'s measure of executive constraint. The measure includes a veto points measure based on economic models of constraint that factor in veto points produced by powerful political parties with different views, congressional and bureaucratic procedures, as well as judicial powers to overrule the executive. The coding scheme for this measure most closely matches theories of institutional constraint and

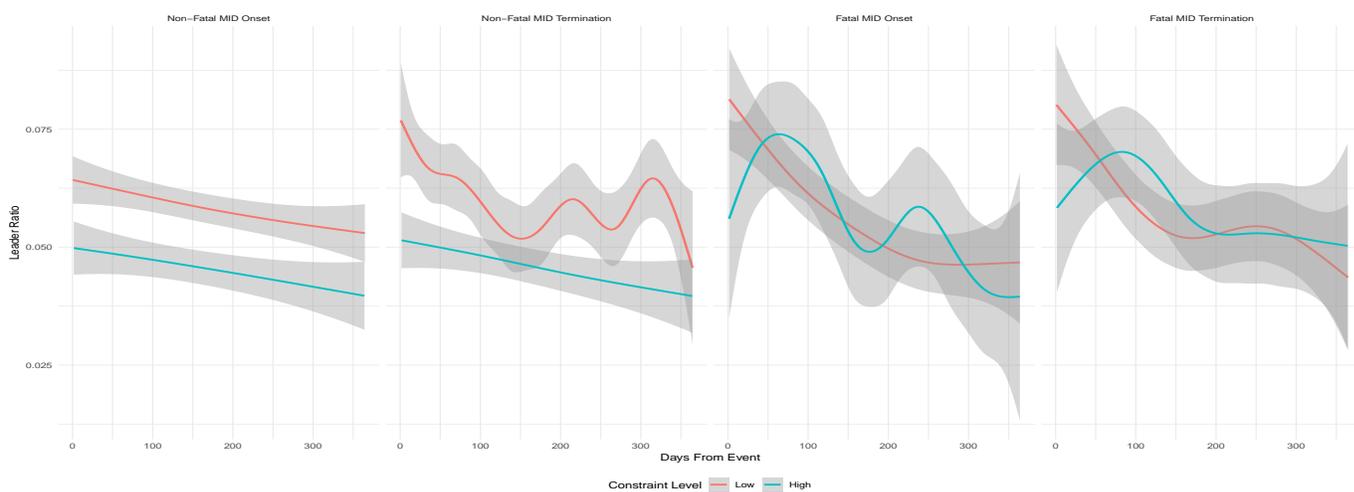
¹¹These measures only cover the last 3 phases of our conflict process. We could not find an analogous measure for crisis instigation.

is commonly used in political economy models (cf [Bütche and Milner, 2008](#)).

To be clear, institutional strength is distinct from regime type. Although many have recognized that democracies are more likely than autocracies to embody constraints and punishments, there are many well documented examples where autocrats can be constrained and punished ([Weeks, 2008](#)). A critical advantage of [Henisz \(2000\)](#)'s measure is that it includes constraints on autocratic regimes. For example, during periods of military rule in Pakistan, the measure captures Pakistan's strong judicial system and thus considers Pakistan slightly constrained.

However, we accept that aspects of the electoral processes inherent in democratic decision-making can influence conflict behavior and also play an important role in selecting leaders. Thus, we also include a measure of regime-type (democracy v. autocracies) that captures the electoral process but does not rely on how that process produces constraints. We use ([Boix et al., 2013](#))'s coding for democracy because this is exactly how they operational their measure (see [Graham et al., 2017](#), for discussion). We are reluctant to use Polity measures because these they include institutional constraint features in their coding.¹² In our extreme bounds analysis (EBA) below we account for many other features of regimes including differences between autocratic governments ([Weeks, 2008](#)).

Figure 5: Leadership Discussion Near Important Stages of Conflict



As a first cut, and to illustrate what a more sophisticated analysis will ultimately do, we re-plot the Lowess lines from Figure 5, but now distinguish between states with an institutional

¹²Polity measures nonetheless do not alter our findings.

constraint score of 0 and those with a positive institutional constraint score.¹³ At the onset of all four events (near day 0 on the x-axis), it seems that the CIA discusses leaders of unconstrained regimes more than leaders of constrained regimes. This provides some preliminary support for the constraint arguments. However, it is unclear if these differences are significant in two of the four cases. For the onset and termination of fatal MIDs, the relative level of discussion for different levels of constraint converge after 50 days. For the onset of and termination of non-fatal MIDs, the discussion does not seem to converge.

3 Statistical Analysis

We want to know how our leadership ratio variable fluctuates given what conflict events are going on in a given week, and also what institutions are influencing a leader’s choices. Below we develop different statistical models to analyze the independent and interactive effects of institutions and conflict. In section 4 we further validate our measure and trace our main findings through a critical case: the Arab-Israeli War (1967).

3.1 Question 1: How important are leaders at different stages of the conflict process on average?

To address this question, we estimate a model:

$$Leader\ ratio_{cw} = \zeta_0 + \alpha_{wc}X_{wc} + \beta_{wc}Z_{wc} + \gamma_{wc}W_{wc} + \delta_c + \epsilon_c \quad (4)$$

The subscripts represent that we measure every variable at the country (c) week (w) level. X is a vector of our conflict event measures, Z is a vector of our institution measures, W is a vector of our shadow variables. ζ_0 is the intercept. We include country level fixed effects (δ_c), and cluster errors at the country level (ϵ_c). By including country fixed effects, we eliminate any variation in the relative importance of different countries. By setting the outcome variable as a ratio of leader mentions over total mentions, we eliminate any effects from an increased interest in the country

¹³About 65% of the sample scores 0. We accept that our cut-off at no constraint (0) versus some constraint is arbitrary. This is a simple example. We draw stronger inferences from our statistical models below that take into account more granular variation in regime constraint.

brought on by events for which we do not control. If a country becomes salient at a specific point in time, then leadership and country discussion should increase.

Table 4: Independent Effects of Conflict Events and Institutions

	DV: Leadership Ratio		
	(1)	(2)	(3)
Crisis Onset	-0.011 p = 0.618	-0.011 p = 0.614	-0.020 p = 0.377
Crisis ends peacefully	0.053* p = 0.017	0.053* p = 0.019	0.055* p = 0.015
Onset of Fat. combat	0.120** p = 0.0002	0.120** p = 0.0003	0.112** p = 0.001
Fat. Combat ongoing	0.0003 p = 0.083	0.0004 p = 0.070	0.0004 p = 0.051
Fat. Combat Ends	0.049 p = 0.120	0.048 p = 0.135	0.039 p = 0.227
Political Constraint		-0.032** p = 0.00000	-0.029** p = 0.00000
Democracy		0.008** p = 0.005	0.008** p = 0.004
Civil War Onset			0.261** p = 0.00001
Fat. Civil War ongoing			0.0004 p = 0.269
Fat. Civil War Ends			0.633** p = 0.000
Observations	109,300	106,943	106,943

Note:

*p<0.05; **p<0.01

All models include country fixed effects and clustered error.

The α coefficients tell us whether proximity to a conflict relevant event predicts increases in relative leader discussion. When the coefficients are positive and significant it implies that the CIA reports on leader-specific activity relative to a country generally more than they would during peacetime, controlling for how important that country is, and other stages of the conflict process. We report the results in Table 4. The first regression only includes our five conflict events. The second adds institutional variables. The third adds shadow measures for conflict events.

We begin with our main measures of the conflict process. Across all three model specifications, we find that the CIA significantly raises leadership discussion at the point of a crisis where leaders must choose between escalating to fatal conflict or finding a peaceful resolution. Our coefficients for the end of fatal combat and ongoing fatal combat are positive in all specifications. However, they are not significant at the .05 level. In some specifications, they are significant at the .1 level. We find no evidence that leaders are more important at the onset of non-fatal crises than they are

in peacetime.

Turning to our shadow variables we see intuitive findings that increase our confidence in the results. First, our shadow conflict variables (from civil war data) follow the same pattern of our main conflict variables (from international conflict data). This supports our interpretation that leaders matter more for certain types of decisions in the conflict process.

Turning to institutional constraints. We find that leaders are less important when constraints are high. While this finding is intuitive for some institutional scholars, it may surprise others given that democracy is correlated with constraints (Wolford, 2007), albeit imperfectly (Weeks, 2008). Critically, we also find that democratic leaders are more important than autocrats on average. This both supports and clarifies differences between those that theorize that regimes constrain and those that theorize that they punish. Leaders are less important on average across all of time when they are constrained heavily by institutions. But once we factor out the constraining features of regimes, leaders of democracies are more important than autocrats, possibly because they respond to punishments.¹⁴

3.1.1 Robustness of results and differential bias: Extreme Bounds Analysis

When there is one theory to test, linear regression is helpful because researchers can specify all the relevant control variables within a single model. We are testing many theories about when and why leaders are important that span several ongoing debates. Each argues that leaders are relevant at different points in the conflict process. Furthermore, they may emphasize how leadership importance is mediated by different institutional configurations. To satisfy these theories, we must simultaneously include and omit events and regime variables from our analysis, as well as other controls about leadership type.

Given there are so many model specifications we could select, it seems especially unfair to report a null finding based on a single regression estimate. We address this concern using an extreme bounds analysis (Leamer, 1985). An EBA is a sensitivity test applied to a regression model. In it, the researcher estimates one regression for every combination of the potentially relevant predictors, and other specification choices (Sala-i Martin, 1997).

First, we identify a list of all the potentially relevant controls. We control for events outside

¹⁴Future work should tease out which of these institutional features are most powerful.

of conflict where past researchers believe leader choices can matter. We control for time until the onset and termination of alliance agreements, as well as entry into the General Agreement on Tariffs and Trade.

We also control for theories that suggest leaders cause uncertainty, and this uncertainty creates opportunities for miscalculation—including research on reputation and the duration of tenure (Croco and Weeks, 2016; Weisiger and Yarhi-Milo, 2015)—by including a measure that capture the key moments of uncertainty about a leader’s discretion: time until a leadership change (including successful coups), time until a failed coup attempt, and time until an election.

We control for other domestic institutional variables that may influence leadership discussion and vary with our measure of constraint. We control for human rights abuses, and how leaders come to power (regularly or irregularly). To make sure that we properly account for nuances in autocratic regimes, we also include one dummy variable for each of the autocratic regime-types suggested by (Weeks, 2008).

To make sure we properly account for variance among leaders, we include all the personal characteristic variables recorded by the LEADs project that have less than 15% missing data.

Second, we estimate many different regressions. We summarize all our regression estimates using the equation:

$$Leader\ ratio_{cw} = \zeta_0 + I_X\alpha_{wc}X_{wc} + I_Z\beta_{wc}Z_{wc} + I_W\gamma_{wc}W_{wc} + i_k\zeta_{wc}K_{wc} + \delta_c + \epsilon_c \quad (5)$$

K_{wc} is a vector of our control variables. I_x, I_z, I_W are vectors of indicator functions of length X, Z, W respectively. We estimate one regression model that includes every combination of relevant conflict events, institutional measures, and shadow variables. We estimate each combination once with the inclusion of all the relevant controls and ($i_k = 1$), and then a second time without the controls ($i_k = 0$). We always include country-level fixed effects and cluster error at the country level. This generates 16,432 unique regression estimates.

Sala-i Martin (1997) recommends making inferences based on the distribution of the direction and significance of coefficients across different regression models.¹⁵ If coefficients are consistently positive (or negative) and significant, then researchers can be more confident that the results do

¹⁵See Levine and Renelt (1992) for other inferential techniques.

not rely on model dependence. [Miller, Joseph, and Ohl \(2016\)](#) argue that researchers can be more confident in a negligible effect when coefficients are reliably close to 0 and insignificant. However, EBA can also call into question a null finding if coefficients are stable across many modeling specification even if they do not reach a significance threshold.

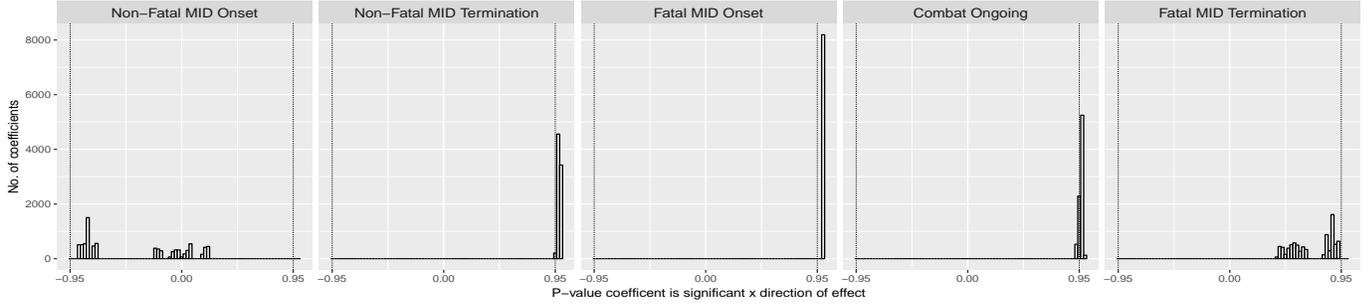
To conclude that leaders are important at a specific stage of a conflict, or given a specific set of institutional features, we hope that our coefficients consistently point in the same direction and are consistently significant. For each regression, we record all of our α , and γ coefficients and their p-values. Each variable (except for country fixed effects) appears in exactly half of the regression models we estimate. For each coefficient we compute: $(1-p\text{-value})(1-2N)$ where N is an indicator function equal to 1 if the coefficient was negative. When this measure is $-.99$, it implies that the regression coefficient was negative and significant using a p-value at $.01$. When this measure is $.95$, it implies that the regression coefficient was positive and significant using a p-value at $.05$. When the measure is close to 0, in either direction, it implies that the coefficient was not significant.

We plot the histogram for every relevant variable across all regressions in [Figure 6](#). The dotted lines mark the density of coefficients that are significant with 95% confidence. The top row plots the histograms of our α coefficients. Two variables are positive and significant in every specification: the onset of fatal MIDs, and the termination of non-fatal MIDs. These are our two measures that capture the point in time where states must choose between escalating a crisis into violent military conflict, and de-escalating a crisis back to peaceful relations. The fact that they are reliably positive and significant suggests leaders are important at this decision node. We also note that in every single specification, the coefficient for termination of fatal MIDs, and ongoing conflict is positive. We take this as weak support that leaders are discussed more during these periods.

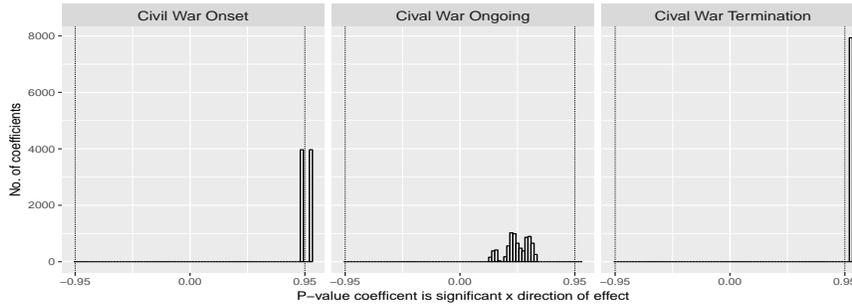
In contrast, in 50% of the regressions that we estimated leaders are discussed less at the onset of a non-fatal MID. Even when the coefficients are positive, their significance is very low. These results suggest that the CIA does not believe leaders are especially important for instigating crises.

We validate our positive findings using our shadow variables plotted in row (b) of [Figure 6](#). Consistent with international conflict, we observe raised leadership discussion at the onset and termination of violent conflict. We could not find support for discussion when conflict was ongoing. These results provide an important validation of our initial findings: at critical escalation and de-escalation junctures leaders seem to matter. However, leaders matter less for getting states into

Figure 6: EBA results



(a) Measures for stage in Conflict Process



(b) Shadow Measures (Civil War)

conflicts. We find mixed support for whether leaders matter while combat is ongoing.

These results also suggest an exciting avenue for future research. We could not find any research on leadership characteristics and civil conflict. But these results suggest that leaders are critical during these periods as well. Future research should consider which characteristics are driving important choices for civil conflict.

Two features of these results increase our confidence that differential biases are not at work. First, our independent variables and controls include 13 different time-varying events. If error in our measures were correlated with different events (especially conflict stages) then our statistically significant parameter estimates, and confidence in them, would fluctuate as we removed and then introduced other events as controls. We do not see this. Rather, there is remarkable consistency across different modeling specifications.

Second, our theory predicts consistent estimates between civil conflict and international conflict outcomes. But this is likely not the case for differential biases. Consider that there may be more redacted text at the onset of an international crises because there is more ongoing diplomacy. But the incentives to redact would be different at the onset of civil war.

3.2 Question 2: How do institutions mediate leadership importance at different stages of the conflict process?

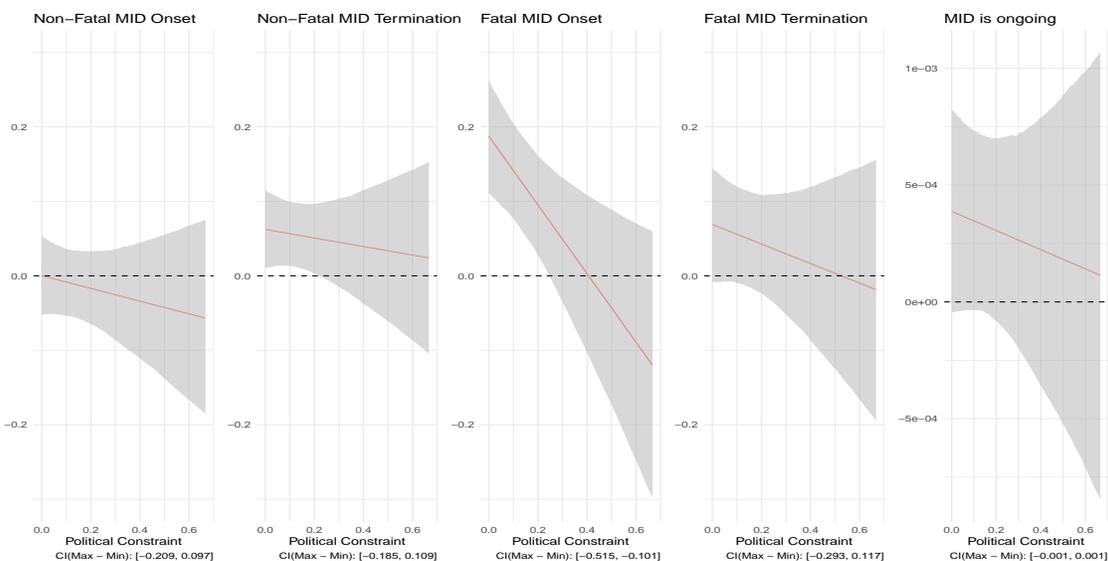
To address this question, we estimate a model:

$$Leader\ ratio_{cw} = \zeta_0 + \psi_{wc} X_{wc} \times Constraint_{wc} + \alpha_{wc} X_{wc} + \beta_{wc} Z_{wc} + \delta_c + \epsilon_c \quad (6)$$

The model includes an interactive term between every stage in the conflict process and our measure of constraints. By combining the interaction term with the constituent terms, we produce four different states of the world based on: (1) whether leaders are constrained by institutions or not, and (2) whether a specific stage of the conflict process is ongoing or not.

We present the key results in Figure 7.¹⁶ These marginal effects plots show our events coefficients as a function of political constraint. Each panel focuses on a single event regression coefficient. The y-axis is the coefficient's effect size. The x-axis is different levels of political constraint. The red solid lines plot the coefficients as a function of political constraint. Looking across all four estimates we note that the marginal effects are always downward sloping. This is broadly consistent with our intuition. Generally speaking, leaders who are unconstrained are more important during conflict events than leaders who are constrained.

Figure 7: Marginal effects for conflict variables as a function of political constraint



¹⁶Results for the full regression are reported in the appendix.

The shaded region reports the 95% confidence intervals that surround the effect size and the dashed lines represent 0. When the shaded region includes 0, it means that at that level of political constraint, we observe no difference between leadership discussion during that specific point in the conflict and at peace-time.

In the baseline model that considered the average effect across all regimes, we found that the CIA thought leaders were important at the point in time where they chose between escalating into violent conflict, or de-escalating to peace. The marginal effects plots show that the results are more nuanced. It seems that the average effect is driven by large increases in leadership importance for unconstrained regimes at these points of the conflict process. However, when leaders are heavily constrained, the CIA is no more likely to discuss them during these periods of the conflict and in peacetime.

In interpreting these marginal effects, we note here and show in the regression results presented in Appendix C that the constituent term for political constraint is also negative and statistically significant. Thus, our complete interpretation of these results is as follows. When a crisis erupts, a leader of an unconstrained regime is more important for how that crisis ends than they are during peacetime. When a crisis erupts, a leader of a constrained regime is no more important for how that crisis ends than they are during peacetime. They are also, on average, less important than unconstrained leaders all points.

In the baseline model we also found weak support for the proposition that leaders are important at the end of fatal MIDs, when wars are terminated. We found no support that leaders were especially important during an ongoing militarized conflict. The marginal effects plots recover leadership importance at these times for minimally constrained regimes. When leaders face minimal constraints¹⁷ the CIA is statistically significantly more likely to discuss them during these periods of the conflict process relative to the country generally. However, the effects become indistinguishable from 0 with even a small amount of political constraint. As a result, we cannot recover an average effect across all regimes.

¹⁷States that score between 0 and 0.1 on our measure include Pakistan and Iran.

4 Illustrative Case: The 6 day War

At around 8:00 AM on June 5, 1967, Israeli pilots launched waves of attacks on Egyptian airfields, devastating the Egyptian air forces and tipping simmering Middle East crisis into full-scale conflict. Thus began the Six Day War, a short but intense conflict that saw Israel annex the Sinai Peninsula, the Gaza Strip, the West Bank, the Golan Heights, and East Jerusalem. The six days of hostilities, the preceding weeks of crisis, and the description of them in the PDB illustrate our findings.

The Six Day War presents a good test of our theory for two reasons. First, there are clear dates for conflict onset and conflict termination. Thus, we can cleanly compare reporting prior to and during hostilities. Second, the crisis and conflict featured leaders who faced different institutional constraints. King Hussein of Jordan and President Gamal Abdel Nasser of Egypt faced limited formal constraints, though they did care for their public standing and were not entirely unencumbered. By contrast, however, Israeli Prime Minister Levi Eshkol faced a highly constrained institutional environment. During this period, the Israeli cabinet and individual ministers acted as a strong check on the prime minister's discretion. As a result, the Six Day War allows us to investigate whether CIA reporting, as we suggest above, varies with a leader's institutional prerogative.

The brewing crisis in the Middle East comes on to the CIA's radar in mid-May, with the PDB reporting on possible retaliation by Israel against Syria following a series of Syrian-sponsored terrorist attacks (PDB, 5/11/1967). The focus quickly turns to Egypt when Nasser escalates the crisis in an attempt to deter Israeli action. According to the CIA, Nasser "threw down the gauntlet and announced the closure of the Gulf of Aqaba to Israeli shipping" on May 22. Nasser ultimately draws Hussein into the crisis, which erupts into conflict on the morning of June 5 when "Israeli planes raided airfields in Cairo and other areas." From May 19 through the termination of the conflict on June 10, the crisis was the top item in every PDB delivered to the president.

During the crisis period, May 11 to June 4, there is considerable CIA reporting on the leaders of the most involved states. The CIA frequently reports its understanding of the leaders' actions, goals, strategies, considerations, constraints, and thought processes. Nasser "continues his gigantic gamble" (PDB, 5/20/1967). Hussein felt compelled to join with Egypt, signing a joint defense pact

in Cairo, “leaving his scalp nailed to Nasir’s lodge pole” (PDB, 5/31/1967).¹⁸ Eshkol “is being hard pressed by the hawks within and outside his government” (PDB, 6/1/1967). Indeed, on May 25 there is an annex to the PDB devoted to Nasser’s behavior and how it “forced the more moderate Arab regimes”—e.g., the Jordanian monarchy—“into a Hobson’s choice.” (PDB, 5/25/1967)

Table 5: Ratio of Leader-to-Country Mentions in the Six Day War

	Nasser	Hussein	Eshkol
Crisis (May 11 - June 4)	0.72 (32/44)	1.16 (14/12)	0.13 (9/67)
War (June 5 - June 10)	0.25 (6/24)	0.11 (1/9)	0.00 (0/33)
Total	0.55 (38/68)	0.71 (15/21)	0.09 (9/100)

Leaders, as our data suggest, are especially important at this particular decision node; leaders decide whether crises fizzle out or else explode into conflict. By contrast, leaders virtually disappear from CIA analysis and reporting during the primary battles of the Six Day War, when they are relatively less important. As Table 5 shows, the ratio of leader mentions to country mentions during the crisis far outstrips the ratio during the war itself for all three leaders. Indeed, during the crucial first three days of fighting, there is only one leader mention. On June 5 and June 7, there are no leader mentions at all. Instead, there are descriptions of military operations. For example, on June 7:

At this point, the shooting continues despite the UN ceasefire resolution. Early this morning Israeli planes were hammering Jordanian positions outside Jerusalem . . . The Israelis appear to hold substantial portions of the Sinai Peninsula, and Cairo is ordering the Egyptian force at Sharm ash-Shaykh on the Straits of Tiran to withdraw. In fact, there are strong indications that the Egyptians may be withdrawing most, if not all, of their forces from the Sinai. (PDB, 6/7/1967)

Only when the outcome of the war—a decisive Israeli victory and Egyptian humiliation—becomes clear, do leaders re-emerge as a focal point of the PDB. In particular, the CIA reports on June 8 that Nasser has personally “informed” other Arab governments that Egypt will not comply with the UN ceasefire. The CIA also reports that “Embassy Cairo believes that public realization of the Arab defeat has generated strong feeling against Nasir, and foreign diplomats in Cairo consider the

¹⁸Contrary to contemporary convention, the CIA uses the spellings “Nasir” for Nasser and “Husayn” for Hussein.

Egyptians are in a state of panic over the military debacle.” (PDB, 6/8/1967) The following day, June 9, the PDB states that “[s]igns are growing that Egypt’s defeat has badly damaged Nasir’s prestige in the Arab world.”

The PDB’s reporting on the Six Day War and the preceding crisis also illustrate some of the more nuanced findings we present quantitatively above. In particular, the ratio of leader to country mentions is lower for Israel than it is for either Jordan or Egypt, the other two principal crisis actors (Table 5). This, our data suggests, reflects institutional differences. King Hussein of Jordan and President Nasser of Egypt were less constrained than their Israeli counterpart. Israel is not only a democracy, but a parliamentary democracy in which leaders are especially constrained. Cabinet politicking was pervasive in Israel during this period. As the PDB would later note, “chronic infighting” was a feature of Israeli politics in the Eshkol years. (PDB, 1/22/1969)

We see evidence of this in the Six Day War. Though Israel is discussed more than either Egypt or Jordan during both the crisis and the war, Eshkol is mentioned less than his less constrained counterparts. A partial explanation for this is that the PDB was not focused only on Eshkol among Israeli leaders. There are frequent mentions of cabinet deliberations. Most interestingly, the CIA is attentive to the role of Moshe Dayan, a general and war hero from the 1956 Suez conflict who is reported to have joined the Israeli cabinet as defense minister on June 1. (PDB, 6/1/1967) It is not just what Eshkol believes and wants that matters; other actors in the Israeli system matter, too.

In addition, we see concern for reputation and prestige in the CIA’s reporting on Nasser. Though Nasser is an autocrat, his influence rested, in the CIA’s view, on his popularity among “the street Arabs of the Middle East,” which was important currency in Egypt and in his dealings with foreign leaders. (PDB, 5/25/1967) If leaders feel that their hands are tied by public or elite opinion, then their choices to escalate crisis may be of particular note. As Nasser escalated the crisis, the PDB reports that while the Egyptian leader was “undoubtedly anxious to duck a fight,” he had “crawled way on the limb. He has made one of he best publicized troop build-ups in recent years and has told the United Nations Emergency Force to go home. With his prestige on he line, Nasir will find it hard indeed to back down if the crunch comes.” (PDB, 5/19/1967)

5 Conclusion

Thanks to the recent proliferation of biographical data on leaders, as well as increasingly precise measures of conflict, we have enormous potential to discover how the first image influences war and peace. Where should we focus our efforts? To answer this question, scholars need to know which leaders are important and when they are important. But a measure of leadership importance has proven elusive. We developed a new cross-national, weekly measure for how important leaders are perceived to be for the conduct of foreign policy.

We show that leaders do not matter equally. Rather leaders who do not face domestic political constraints are systematically seen as more important at many stages of the conflict process than they are during peacetime. These leaders play a large role in negotiating how crises end, and also play moderately important roles during conflict and at the termination of conflict relative to the role they play in peacetime. However, leaders who face severe political constraints are seen as less important than unconstrained leaders on average. They are also no more important during conflict than they are at any other time in their tenure.

These results will help leadership scholars theorize the mechanisms that link specific characteristics to conflict outcomes, and also model the complex inter-connected stages of conflict and diplomacy. They also highlight that political constraints and other features of regimes (such as punishments) play different roles that must be accounted for in theoretical and empirical models. Finally, leader biography scholars can use our measure of importance to address endogeneity concerns they face when trying to link characteristics to conflict outcomes.

Our findings have important policy implications. At the international level, policymakers may temper their demands to avoid conflict when they face a new, and untested, foreign leader. Our findings suggest that leaders are most salient for state decision-making during pre-conflict, militarized crisis periods. Thus, it is reasonable to temper demands in a crisis when facing an unknown foreign leader. However, American policymakers need not to temper their demands as much during peace-time or war-time negotiations because leadership variables matter less at these times. At these times, they should bargain just as forcefully as they would have if they had a longer history with a foreign leader. For a better known leader, our theory suggests that if a foreign leader is pragmatic, it is desirable to negotiate a settlement at the point of escalation because this is the

point that this leader has the most discretion.

At the domestic level, states should look beyond regime type to forecast which domestic actors will be important for specific negotiations during conflict. The power brokers who negotiate how conflict unfolds will vary depending on the phase of conflict a state is in, and how specific institutional configurations impact the state's decision-making process.

At the bureaucratic level, we identify a new way that the CIA can exploit data science to improve intelligence evaluation. The CIA is so confident that data science is essential for analysis that it raised digital innovation to the Directorate level (Lowenthal, 2019). However it has faced challenges integrating insights from average effects (which statistics provide) to specific cases (which analysts at the Directorate of Analysis typically focus on) (Jervis, 2010). At the same time, the CIA has searched for ways to better evaluate their products without tracking the production of individual analysts (Marrin, 2012).¹⁹ This paper highlights that textual analysis can search for patterns in existing intelligence products. By monitoring variation in these patterns, the CIA can better understand regularities in their reporting procedures.

¹⁹See also <https://foreignpolicy.com/2016/05/04/measuring-change-at-the-cia/>

References

- Bak, D. and G. Palmer (2010, jul). Testing the Biden Hypotheses: Leader Tenure, Age, and International Conflict. *Foreign Policy Analysis* 6(3), 257–273.
- Besley, T., J. G. Montalvo, and M. Reynal-Querol (2011). Do educated leaders matter? *The Economic Journal* 121(554), F205–227.
- Blight, J. G. and D. A. Welch (1998, sep). The Cuban missile crisis and intelligence performance. *Intelligence and National Security* 13(3), 173–217.
- Boix, C., M. Miller, and S. Rosato (2013, dec). A Complete Data Set of Political Regimes, 1800–2007. *Comparative Political Studies* 46(12), 1523–1554.
- Büthe, T. and H. Milner (2008, oct). The politics of foreign direct investment into developing countries: increasing FDI through international trade agreements? *American Journal of Political Science* 52(4), 741–762.
- Byman, D. L. and K. M. Pollack (2001, apr). Let Us Now Praise Great Men: Bringing the Statesman Back In. *International Security* 25(4), 107–146.
- Chiozza, G. and A. Choi (2003, jun). Guess Who Did What: Political Leaders and the Management of Territorial Disputes, 1950–1990. *Journal of Conflict Resolution* 47(3), 251–278.
- Chiozza, G. and H. E. Goemans (2011). *Leaders and international conflict*. Cambridge University Press.
- CIA (2008, apr). DCI Walter Bedell Smith Creates the Directorate for Intelligence. *Featured Story Archive*.
- Croco, S. E. and J. L. P. Weeks (2016, oct). War Outcomes and Leader Tenure. *World Politics* 68(4), 577–607.
- Davis, J. (1995). A Policymaker’s Perspective on Intelligence Analysis. *Studies in Intelligence* 38(5).
- Downs, G. W. and D. M. Roche (1994, may). Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War. *American Journal of Political Science* 38(2), 362.
- Dyson, S. B. (2009). “stuff happens”: Donald rumsfeld and the iraq war. *Foreign Policy Analysis* 5(4), 327–347.
- Dyson, S. B. and C. A. Duelfer (2020, may). Assessing How the U.S. Intelligence Community Analyzes Foreign Leaders. *International Journal of Intelligence and CounterIntelligence*, 1–29.
- Ellis, C. M., M. C. Horowitz, and A. C. Stam (2015, aug). Introducing the LEAD Data Set. *International Interactions* 41(4), 718–741.
- Fearon, J. D. (1994, jun). Signaling Versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model. *Journal of Conflict Resolution* 38(2), 236–269.
- Friedman, J. A. (2019). *War and chance : assessing uncertainty in international politics*. Oxford University Press.

- Fuhrmann, M. (2020, jan). When Do Leaders Free-Ride? Business Experience and Contributions to Collective Defense. *American Journal of Political Science*, ajps.12502.
- Fuhrmann, M. and M. C. Horowitz (2015). When leaders matter: Rebel experience and nuclear proliferation. *The Journal of Politics* 77(1), 72–87.
- George, R. Z. and J. B. Bruce (2008). *Analyzing intelligence : origins, obstacles, and innovations*. Georgetown University.
- Gibler, D. M., S. V. Miller, and E. K. Little (2016, dec). An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001. *International Studies Quarterly* 60(4), 719–730.
- Gift, T. and D. Kremaric (2017, mar). Who Democratizes? Western-educated Leaders and Regime Transitions. *Journal of Conflict Resolution* 61(3), 671–701.
- Goemans, H. (2000). *War and punishment : the causes of war termination and the First World War*. Princeton studies in international history and politics. Princeton, N.J: Princeton University Press.
- Goemans, H., K. Gleditsch, and G. Chiozza (2009). Introducing Archigos: A Dataset of Political Leaders. *Journal of Peace Research* 46(2), 269–283.
- Gottfried, M. S. and R. F. Trager (2016, jun). A Preference for War: How Fairness and Rhetoric Influence Leadership Incentives in Crises. *International Studies Quarterly* 60(2), 243–257.
- Graham, B., M. K. Miller, and K. W. Strom (2017, nov). Safeguarding Democracy: Powersharing and Democratic Survival. *American Political Science Review* 111(4), 686–704.
- Grynaviski, E. (2018). *America’s middlemen : power at the edge of empire*.
- Henisz, W. J. (2000). The Institutional Environment for Multinational Investment. *Journal of Law, Economics, and Organization* 16(2), 334–64.
- Heuer, R. J. (1999). *Psychology of Intelligence Analysis*. Lulu.com.
- Horowitz, M. C. and M. Fuhrmann (2018, nov). Studying Leaders and Military Conflict. *Journal of Conflict Resolution* 62(10), 2072–2086.
- Jervis, R. (2010). *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*. Cornell Studies in Security Affairs. Cornell University Press.
- Jervis, R. (2013, apr). Do Leaders Matter and How Would We Know? *Security Studies* 22(2), 153–179.
- Kaelberer, M. (1997, mar). Hegemony, Dominance or Leadership?: Explaining Germany’s Role in European Monetary Cooperation. *European Journal of International Relations* 3(1), 35–60.
- Katagiri, A. and E. Min (2019, feb). The Credibility of Public and Private Signals: A Document-Based Approach. *American Political Science Review* 113(1), 156–172.
- Kertzer, J. D. (2016). *Resolve in international politics*. Princeton University Press.
- Koch, M. T. and S. A. Fulton (2011). In the defense of women: Gender, office holding, and national security policy in established democracies. *The Journal of politics* 73(1), 1–16.

- Kovar, R. (2000). Mr. Current Intelligence: An Interview with Richard Lehman. *Studies in Intelligence* 1(9), S51–S63.
- Krcmaric, D. (2018, apr). Should I Stay or Should I Go? Leaders, Exile, and the Dilemmas of International Justice. *American Journal of Political Science* 62(2), 486–498.
- Krcmaric, D., S. C. Nelson, and A. Roberts (2020, may). Studying Leaders and Elites: The Personal Biography Approach. *Annual Review of Political Science* 23(1), annurev-polisci-050718-032801.
- Leamer, E. (1985). Sensitivity Analyses Would Help. *American Economic Review* 75(3), 308–13.
- Levine, R. and D. Renelt (1992). A Sensitivity Analysis of Cross-Country Growth Regressions. *American Economic Review* 82(4), 942–963.
- Lindsey, D. (2017, sep). Diplomacy Through Agents. *International Studies Quarterly* 61(3), 544–556.
- Lindsey, D. (2019, oct). Mutual Optimism and Costly Conflict: The Case of Naval Battles in the Age of Sail. *The Journal of Politics* 81(4), 1181–1196.
- Lindsey, D. and W. Hobbs (2015, oct). Presidential Effort and International Outcomes: Evidence for an Executive Bottleneck. *The Journal of Politics* 77(4), 1089–1102.
- Lowenthal, M. M. (2019). *Intelligence : from secrets to policy*.
- Mandel, D. R. and A. Barnes (2014, jul). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences* 111(30), 10984–10989.
- Mandel, D. R. and A. Barnes (2018, jan). Geopolitical Forecasting Skill in Strategic Intelligence. *Journal of Behavioral Decision Making* 31(1), 127–137.
- Mandel, D. R. and P. E. Tetlock (2018, dec). Correcting Judgment Correctives in National Security Intelligence. *Frontiers in Psychology* 9.
- Mansfield, E. D., H. V. Milner, and B. P. Rosendorff (2002, aug). Why Democracies Cooperate More: Electoral Control and International Trade Agreements. *International Organization* 56(3), 477–513.
- Marrin, S. (2012, dec). Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure? *Intelligence and National Security* 27(6), 896–912.
- Marshall, M., K. Jaggers, and T. Gurr (2002). Polity IV dataset.
- McGillivray, F. and A. Smith (2006, may). Credibility in Compliance and Punishment: Leader Specific Punishments and Credibility. *The Journal of Politics* 68(2), 248–258.
- McManus, R. W. (2019, dec). Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion? *British Journal of Political Science*, 1–19.
- Mearsheimer, J. J. (2001). *The Tragedy of Great Power Politics*. New York: Norton.
- Mesquita, B. B. D., A. Smith, and R. M. Siverson (2005). *The logic of Political Survival*. MIT Press.

- Miller, M. K., M. Joseph, and D. Ohl (2016, may). Are Coups Really Contagious? An Extreme Bounds Analysis of Political Diffusion. *Journal of Conflict Resolution*, 0022002716649232.
- Milner, H. V. and D. H. Tingley (2015). *Sailing the water's edge : the domestic politics of American foreign policy*. Princeton University Press.
- Min, E. (2021). Talking while Fighting: Understanding the Role of Wartime Negotiation. *International Organization* 2.
- Palmer, G., V. D'Orazio, M. R. Kenwick, and R. W. McManus (2020, jun). Updating the Militarized Interstate Dispute Data: A Response to Gibler, Miller, and Little. *International Studies Quarterly* 64(2), 469–475.
- Polsky, A. (2013). *Elusive Victories: The American Presidency at War*. Oxford University Press.
- Priess, D. (2016). *The president's book of secrets : the untold story of intelligence briefings to America's presidents*.
- Reiter, D. and A. C. Stam (2002). *Democracies at war*. Princeton University Press.
- Robarge, D. S. (2005). Getting It Right CIA Analysis of the 1967 Arab-Israeli War. *Studies in Intelligence* 49(1).
- Sala-i Martin, X. (1997). I Just Ran Two Million Regressions. *American Economic Review* 87(2), 178–183.
- Saunders, E. (2011). *Leaders at war : how presidents shape military interventions*. Cornell University Press.
- Saunders, E. N. (2017, apr). No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making. *International Organization* 71(S1), S219–S247.
- Schelling, T. C. (1957, mar). Bargaining, Communication, and Limited war. *Journal of Conflict Resolution* 1(1), 19–36.
- Trager, R. F. (2010, may). Diplomatic Calculus in Anarchy: How Communication Matters. *American Political Science Review* 104(02), 347–368.
- Waltz, K. N. (2001). *Man, the state, and war: a theoretical analysis*. New York: Columbia University Press.
- Weeks, J. L. (2008). Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62(01), 35–64.
- Weisiger, A. and K. Yarhi-Milo (2015, may). Revisiting Reputation: How Past Actions Matter in International Politics. *International Organization* 69(2), 473–495.
- Wolford, S. (2007, oct). The Turnover Trap: New Leaders, Reputation, and International Conflict. *American Journal of Political Science* 51(4), 772–788.
- Wu, C. X. and S. Wolford (2018, nov). Leaders, States, and Reputations. *Journal of Conflict Resolution* 62(10), 2087–2117.
- Yarhi-Milo, K. (2014). *Knowing the adversary : leaders, intelligence, and assessment of intentions in international relations*. Princeton University Press.

Appendix

Table of Contents

A Coding Procedure	1
A.1 Constructing CORPUS	1
A.2 Converting the photo-copied text into machine readable text	1
A.3 Regular expression	2
A.4 Iterative Dictionary Construction	2
A.5 Why leaders, countries, and capitol cities, but not other terms?	4
B Organization Chart	4
C Interaction of Conflict Events and Institutions	5

A Coding Procedure

In this section we document our procedure for converting the PDB text into data.

A.1 Constructing CORPUS

The PDBs are made available through the CREST archive; hosted online on the CIA website. CREST organizes the documents into two libraries based on the years published. The first library includes PDBs 1961-1969. The second includes PDBs 1969-1977.

Each PDB is assigned its own web-page. Each webpage provides a link to a PDF photocopy of the original document, and associated metadata. The most relevant fields in this metadata were (1) the publication date, (2) the number of PDBs published that day.

Our initial scraping effort extracted one record for every webpage and stored it in an SQL database. That record contained one field for each meta-data item, and one field that contained the PDB.

A.2 Converting the photo-copied text into machine readable text

The PDFs are photocopies of the original PDBs. They are not text readable. Thus, our next step was to convert each PDF picture into readable text.

To do this, we used the professional version of Abby Finereader v14 (Abby) made available to us through the [Removed for review] library. We also took advantage of library staff who specialize in text recognition programming. These library scientists helped us tailor the Abby program to extract an accurate record of text.

The photocopies were sometimes photocopied crooked, and each document had a header and footer with labeling we wanted to remove. To address these issues, we first passed the documents through a python script that re-orientated the documents so the lines were straight, and identified and removed headers and footers from the document. With the documents aligned, we entered each PDF as a separate file into Abby.

By default, Abby treats each page as a single column of text. It then identifies text rows, and convert each horizontal row into a line of text. This creates problems for us because several pages of the PDB are organized into columns. If we want to keep the text in order, we need Abby to recognize column breaks and apply the OCR algorithm to each column individually.

Alone, the column structure is not an issue for us because we are counting words that could appear in any order in the document. However, the PDB often breaks up the last word on a line with a hyphen. If the last word on the line is a country, leader or city name, then we introduce error if we do not keep the columns separate.

We tried a variety of automated approaches to recognizing column partitions. However, all were prone to error. For example, we tried to recognize vertical partitions through estimating white pixilated space. However, the inclusion of redaction codes at odd parts of the document forced the algorithm to place column breaks in strange places.

Abby allows the researcher to manually specify column breaks on pages. Unfortunately, the document formats varied to some degree. For example, some PDBs do not introduce a column structure. Others included two evenly spaced columns. Others still included the paragraph's title in a short, left-hand column; then the main text in a wider column on the right. A handful of entries were formatted as double-space all caps.

In addition, some PDBs included a main report and an Appendix section. These two sections were often formatted differently. In each case, the relevant text in the appendix was formatted as a

single page. We wanted to keep the information in the Appendix for our analysis (we ran a version without the Appendix and found no difference in our analysis).

In the end, we developed 4 templates to manage the different structures. We then visually inspected each document and manually applied the templates. This procedure ensured that the text was read in the correct order.

Abby also auto-identifies pictures and treats them separately to text (even when we apply a template). This was helpful for us because the PDB includes maps and photographs that occasionally have text on them. We explicitly did not want to read the text on maps. The reason was that the maps regularly labeled different countries from an entire region but were there to support one entry on a single country (e.g. a map of Vietnam would have Laos and Cambodia on it. But these other countries were not relevant in the analysis). As a result, we told Abby to skip pictures.

Once we had the documents covered, we applied the OCR algorithm to the document. This turned the PDF into machine readable text.

Next, we searched for errors in the conversion process. A nice feature of Abby is that it identifies characters that the program cannot confidently estimate as a specific character. It allows the user to specify what that letter is. Most of the mis-estimations related to punctuation. For example, Abby would read black marks on the page as a comma. We corrected these where we saw them. However, since our objective was to extract words from the entire document, we did not overly focus on punctuation. We then spot-checked 40 PDBs and looked for errors in character recognition of letters. We found two cases where Abby had mis-estimated letters (one e and one g). With such a low error rate, we moved on.

After making these changes, we extracted the text from the documents as .txt files.

A.3 Regular expression

Our next step was to clean the processed text via regular expression code. Our first task was to record, then clean out the codes for redacted text in the document. Our second task was to identify disconnected words and re-connected them over lines. We then performed a variety of other cleaning procedures including removing white spaces, tabs, and certain punctuation.

We spot checked 100 entries against the original PDB. In many cases, the title for an Entry (e.g. Sino-American relations) was not above the relevant entry. But we found no errors that would affect our dictionary guided approach.

A.4 Iterative Dictionary Construction

Our goal was to develop three distinct dictionaries for: countries, capital cities and leaders. We used the ARCHIGOS dataset to determine the list of countries we would focus on, and the list of leaders in charge of those countries on any given day. We used the CIA World Factbook to determine the Capital Cities of all countries in the ARCHIGOS Dataset. When there was a difference between the symbolic (e.g. Jerusalem) and governance capitol (Tel Aviv), we focused on the Governance Capitol.

The PDB is extensively edited before it reaches the president. Further, the CIA standardizes spellings for countries, leaders and capitol cities. This made spelling errors in the documents uncommon. This is not to say we found no spelling errors or variation in spelling. Only that they were relatively rare. This made our job easier, but there were still some significant challenges we faced along the way. The construction of each of the three dictionaries presented their own challenges. We addressed these challenges as they arose.

We started with a construction of the country and capital city dictionaries. Our initial list of spellings came from the CIA World Factbook, ARCHIGOS and the ICEWs dictionary for locations and political actors. One draw back of the ICEWs dictionary included a set of search terms that was too broad (including colloquial names for countries that shared a spelling with common usage words. However, one advantage of ICEWs was that it included a longer list of spellings that covered different parts of speech. For example, ICEWs includes China, Chinese, PRC and other variants of China that all fall within our understanding of country references that do not apply to the leader. Using the different spellings was helpful because we could introduce a long list of terms to search for specifically, rather than use a word stem approach such as “Chin*” that would return too many false positives. Notably, ICEWs includes the country-name conventions “Sino” as a spelling for China, and “Anglo” as a spelling for England. These feature extensively in the PDB near hyphens. Thus, we could introduced the root words, and searches that included hyphenated versions such as “Sino-*”, “.-Sino” and related terms into our search.

One challenge with cities (and some countries) was that some city names included multiple tokens (e.g. Ho Chi Min City). In some cases, we could deal with this by focusing on a distinctive word in the city name. But in other cases, it was unclear if a single token would return false positives. There were not many double token cities. So we manually searched the PDB to determine how the CIA described those cities. We dealt with this issue by entering a double (or sometimes triple) token into the dictionary. We then pre-processed the text to identify these double-tokens before we tokenized the text.

As a first cut, we token-ized the text at the document level, then ran both country and city dictionaries and analyzed results at the country level. Following We identified cases where (1) countries were never mentioned; (2) countries were mentioned but cities were not; (3) countries or cities were mentioned an unusually large number of times. In each of these cases we manually checked the dictionary’s performance and adjusted.

As a second cut, we ran the revised dictionary and analyzed results at the document-country level. We looked for cases where (1) the city was mentioned but the country was not; (2) the city was mentioned more than the country; (3) the document included an unusually small (or large) number of different cities and countries. We then reviewed these cases, revised.

Next we constructed a leader dictionary. Unlike the other two, leaders are not constant over time. Thus, we could not construct a dictionary using country as the unit of analysis. Instead, we developed leader-level dictionary. We started with spellings from Archigos, and the CIA World Factbook.

However, these sources were less effective for leaders than they were for countries. The main reason is that all leaders have more than one name (first and last name at a minimum) and it was unclear what names the CIA used. For example, for certain Asian and Latin American states the CIA referred to leaders by their first name. In addition, transliteration conventions have changed over time, particularly for Arabic names (e.g., Hussein vs. Husayn or Nasser vs. Nasir). As part of another project, we had a person coding large portions of the PDB. We asked that person to identify leader spellings and keep them in a spread sheet. That human coding effort helped considerably.

We then ran diagnostics using the similar techniques as in the country-city case, such as looking for cases of no leader mentions or very few leader mentions relative to country and capital city mentions and updated our leader dictionary as necessary. We discovered two additional challenges. First, there were a handful of cases where two leaders with the same name ruled different countries at the same time (e.g., individuals named Castro led Ecuador and Cuba at the same time). Second, some leaders had common usage names (Park, Price) that came up in many other contexts. Fortunately, the PDB almost always titles each entry with a country name. Thus, there should almost never be a case where there are 0 country mentions and a leader mention.

Using this insight, we focused our leadership count to cases where the country or capital city were also mentioned at some point in the document. In other words, if the country and city count equalled 0, then we assumed the leader count also equalled 0. We then validated that this technique helped alleviate our main concern, and did not affect the counts of other leaders. As part of this process, we identified a handful of additional country and city spellings. We went back and adjusted the dictionary where appropriate. There was one case (leaders named Sanchez over a 2-year period) where the technique counted leaders incorrectly. In this case, we resolved to hand-count these cases.

With our three dictionaries in place, we then entered a validation phase. In the validation phase, we ran the three dictionaries over the tokenized text, extracted counts, and applied the 0 country + city = 0 leader rule. We then randomly chose 20 PDBs and manually counted each country, city and leader mention in the original PDF document. We matched that count to the dictionary count. We repeated this 3 times. In each audit, the dictionary accuracy was over 95%. Thus, we concluded that the dictionary was a comprehensive and reasonable counting procedure.

A.5 Why leaders, countries, and capitol cities, but not other terms?

We omit pro-nouns largely for practical reasons. However, we do not think this damages our inferences for three reasons. First, the PDB uses pro-nouns to refer to leaders, cities and countries. It is common, for the PDB to use “they” or “it” as a pronoun for a city (depending on the context). Thus, our omission of pro-nouns for leaders also reduces the count of city and country mentions when they are used as actors. Second, we cannot think of any reason that the use of pro-nouns over names would be correlated with conflict. As we discuss in the manuscript, this is critical for damaging our inferences. Third, our review of the PDB’s language patterns finds that the PDB often refers to leaders repeatedly in the same paragraph. Thus, their use of pro-nouns is limited to some degree, allowing for multiple mentions of leaders where appropriate (usually because they are discussing a leader’s actions in addition to some other actor’s action and they need to refer to the different actors by name to keep the meaning clear).

We also omit lower-level officials for two theoretical reasons. First, it is unclear who the complete list of actors would be. Should we count only foreign ministers? Or should we count all cabinet officials, business people with close ties to the regime, etc.? There is no obvious rule because the relevant alternative actors will vary depending on the political context of a specific country.

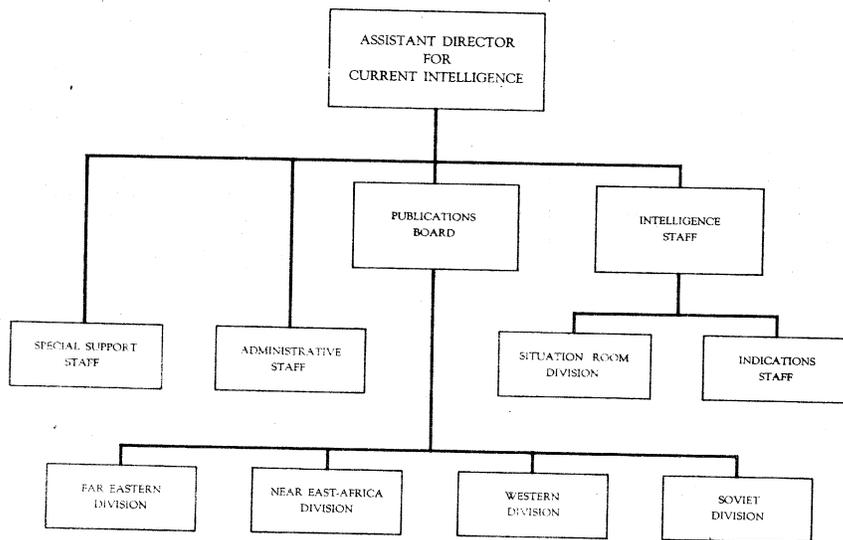
Second, we believe that including other officials would artificially deflate the importance of leaders in periods where the CIA judges the leader’s choices to be important. The reason is that when the CIA describes the actions that a leader is taking, they also discuss how the leader delegates to other staff, or how the leader overcomes debates with other political actors. In cases like this, the CIA clearly believes that the leader is an important actor (and mentions her by name), but also mentions other actors as well. Ultimately, the way that we count is an empirical question, and we validate our strategy using a survey that we discuss in the manuscript. We explain how we aggregate the measure in the manuscript.

B Organization Chart

Below is the organization chart for the CIA’s Office of Current Intelligence, responsible for producing the President’s Daily Brief.

SECRET
Security Information
OFFICE OF CURRENT INTELLIGENCE

25X1A
[REDACTED]
Figure 51
20 March 1947



SECRET
HS/HC-147

C Interaction of Conflict Events and Institutions

Table 6 reports regression results for our model interacting conflict stages and our measures of political constraints (see Section 3.2). Consistent with our baseline model, we find that political constraints are a robust negative predictor of leadership importance. This implies that on average leaders of constrained regimes are discussed less than leaders of unconstrained regimes across all

periods in our dataset.

Table 6: Independent Effects of Conflict Events and Institutions

DV: Leadership Ratio	
Crisis Onset	0.0002 p = 0.996
Crisis ends peacefully	0.062* p = 0.020
Onset of Fat. combat	0.187** p = 0.00001
Fat. Combat ongoing	0.0004 p = 0.085
Fat. Combat Ends	0.068 p = 0.076
Political Constraint	-0.028** p = 0.00001
Democracy	0.008** p = 0.006
<hr/>	
Political Constraint ×	
Crisis Onset	-0.084 p = 0.464
Crisis ends peacefully	-0.059 p = 0.607
Onset of Fat. combat	-0.460** p = 0.004
Fat. Combat ongoing	-0.0004 p = 0.631
Fat. Combat Ends	-0.129 p = 0.404
Observations	106,943
<i>Note:</i>	*p<0.05; **p<0.01