Problems of Warning: Forecast Precision and Forecast Accuracy in Declassified State Department Cables

Great Power Military Intervention Fall 2015 Final Paper

Andrew Halterman

16 December 2015

Puzzle

Why is there relatively little early action or preventative diplomacy to stop or mitigate civil war onset or mass atrocities? Researchers studying the ability of interveners to prevent or mitigate civil war and atrocities rightfully emphasize the importance of early warning for effective responses (Jentleson 2000; Sambanis 2002; Kuperman 2004), yet little is actually observed. Because the United States government ends up in costly interventions in humanitarian conflicts, any theory of international relations or foreign policy that assumes a degree of rational planning in U.S. foreign policy must believe that the United States government has a strong incentive to seek out early warning signals and intervene early in conflicts to avoid more difficult and costly interventions later.

Layout of the Paper

This paper discusses several explanations for this lack of early action before arguing that a dearth of early action is best explained by the limitations of intelligence. I then advance an ar-

gument about the proper way to conduct retrospective analysis of intelligence and intelligence failures, before reviewing the existing literature on early warning, intelligence, and preventive action. In the empirical section of the paper I describe the dataset I use to examine my theory about the link between intelligence and early action. This dataset is a comprehensive collection of 1.8 million declassified U.S. State Department cables sent between 1972 and 1978. Finally, I describe a machine-assisted text analysis approach to searching for early warning in the corpus, and conclude with the limitations of my study and what I conclude about early warning.

Why So Little Intervention?

Military interventions in humanitarian crises (civil wars, conflict-exacerbated famines, and mass killing) have become a common and costly component of U.S. use of power. Both scholars and practitioners argue that intervening early in crises can dramatically reduce the difficulty and cost of intervention (Jentleson 2000; Kuperman 2004; Sambanis 2002). The call for early action or preventative diplomacy came early after the dissolution of the Soviet Union from the United Nations, which was feeling the strain of greater peacekeeping operations, and wanted to forestall constant peacekeeping deployments wherever possible (Boutros-Ghali 1992). Jentleson finds several cases, including Macedonia, North Korea, and Congo-Brazzaville where preventative diplomacy kept the countries from sliding into civil war, saving civilian lives as well as forestalling future demands for the United States or other rich countries to intervene to stop the wars after the began (2000). He finds strong reasons to engage in preventative diplomacy and preventative action. Skeptics about its actual efficacy in practice (Stedman 1995).

Despite the appeal of early or preventative action, we see only a small number of committed preventative actions early in conflicts (with the 2001 deployment of peacekeepers to

Macedonia in 2001 and the U.S. effort to secure the Ohrid Agreement perhaps the most significant since 1989). If the costs to intervene early are so much less than later involvement, why do we see so little? Three explanations present themselves.

The first argument for why early action is uncommon points to U.S. strategic interests. Countries experiencing civil wars are often peripheral to core US interests, and their wars may not warrant even minimal expenditures of U.S. power to mitigate (Posen 2014). The United States often does intervene in these conflicts at later and more costly points, but a defender of this position could argue that as conflicts become worse, they change the strategic calculus for the United States in ways that were not foreseeable at the beginning of the conflict.

A second argument for why early action is rare is that domestic political processes make early action infeasible. In this account, the limiting factor in early action is in persuading Congress or the executive that the American public would stomach or at a minimum not oppose expending U.S. power in a place without strong security interests (as humanitarian interventions are almost by definition) (Western 2005). This persuasion becomes even more difficult when the conflict is in its earliest stages and when there are not ongoing atrocities to point to as justification. Furthermore, successful interventions will prevent atrocities or war from happening in the first place, depriving interveners of the very images that would justify humanitarian interventions in ongoing conflicts, even if early intervention is easier or cheaper than later intervention. The paradox of the prevention counterfactual leaves interveners open for criticism that the atrocities they feared would happen would in fact not have (Kuperman 2013).

The third potential account of why policymakers do not intervene early argues that cognitive biases are responsible for the low number of early interventions. Humans have a well-documented tendency to discount unpleasant information or new data that contradicts previous beliefs about a subject or that complicates otherwise neat narratives (Kahneman 2011).

People also have difficulty separating relevant and irrelevant information for making good decisions, and may ignore important information that arrives slowly in piecemeal format:

> Actors can more easily assimilate into their established image of another actor information contradicting that image if the information is transmitted and considered bit by bit than if it comes all at once. In the former case, each piece of discrepant data can be coped with as it arrives and each of the conflicts with the prevailing view will be small enough to go unnoticed, to be dismissed as unimportant,or to necessitate at most a slight modification of the image (e.g., addition of exceptions to the rule). When the information arrives in a block, the contradiction between it and the prevailing view is apt to be much clearer and the probability of major reorganization will be higher (Jervis 1968, 465–66).

Finally, decision makers with uncertain information find it easier to wait for more intelligence to arrive rather than to make a decisive judgement to act (Chan 1979). These are all characteristics of early warning and intelligence reporting, making response to early warning an inherently psychologically difficult task. This account of the failure of early action would hold that policymakers simply do not have the right cognitive capacity in their situation to make an early decision to intervene.

All of these explanations assume, however, that that knowledge of impending civil war or genocide is known at the time, and that a lack of political will (which could refer to any of the three explanations I provide above) is responsible for a failure to act. This argument of a lack of political will in the face of overwhelming early evidence is a staple in many accounts of US failure to act in the face of genocide or civil war (Feil 1998; Stanton 2012; Power 2013). The paradigmatic example of what Alexander George calls the "warning-response problem" (1997) for genocide is the case of the Rwandan genocide and Dallaire's so-called "genocide fax" (Gourevitch 1998), which relayed the report of an informant in Rwanda who described a plan to "provoke a civil war" and to kill a number of Belgian peacekeepers to force their withdrawal. The existence of this fax, some argue, makes it clear that the Rwandan genocide

was foreseeable at the time, and that failure to take preventative or palliative action must be attributable to something besides poor information.

Similarly, in Somalia, Lyons and Samatar (1995) argue that "concerted diplomatic initiatives at a relatively modest cost in the late 1980s, however, could have prevented the chaos and humanitarian crisis that precipitated the costly military intervention in late 1992" (26), the lack of which is puzzling given that, according to the authors, "International actors, most notably the United States [...] recognized the warning signs of disaster but lacked the interest and the strategy to act effectively" (25). It is this argument, that the failure to act was a failure of will, that I dispute here.

Hypothesis

I argue that the lack of early action/preventative diplomacy in cases of civil war can be explained by a lack of early warning. Specifically, I hypothesize that early warning happens, but it happens so much that "over-warning" prevents policymakers from paying attention. That is, the precision is low, so warnings have relatively low informational content and are discounted. If we find over-warning in less-classified settings, we would have a strong sense that over-warning happens everywhere.

There are two mechanisms through which intelligence could fail and thus preclude early action, each way with different observable implications. The first and obvious way that poor intelligence could limit early action is in a failure to foresee the conflict or atrocity. Without advance warning of a civil war onset or atrocity, policymakers cannot act early to prevent. Evidence for this point should be clear in archival records. Evidence of early warning will exist, and will likely be pointed out by contemporaries.

More pernicious than the failure to warn is over-warning. Hypervigilance by intelligence agencies in looking for warning signs of civil war can produce over-warning, reducing poli-

cymakers' trust in warning reports to the point that they do not gain any extra information from the warning report. The evidence of this should also be clear in archived text, though more difficult to find. Evidence may exist that early warning existed where it should (i.e. before actual outbreaks of war or mass killing), but it would also exist where it should not, in countries that did not experience civil war onset.

If early warning is found everywhere, including where it should not be, this would lend evidence to the second interpretation, that over-warning is partially responsible for a failure to act. If no early warning is found, that would support the argument that policymakers do not act early because they do not know they should. If early warning is found only in the correct times and about the right places, that would suggest that my hypothesis that a lack of warning explains the lack of action is flawed, and that one of the three mechanisms I discuss above are likely to be the better accounts.

How to study intelligence

Retrospective analysis of intelligence failures is fundamentally hobbled by hindsight and availability biases and problems of case selection (Chan 1979). If unmitigated, these biases and case selection processes can prevent researchers from understanding intelligence processes. Intelligence postmortems often reconstruct a series of events that in hindsight make an event seem obvious, even when analysts and readers try to limit themselves to what was knowable at the time. This retrospective bias is almost unavoidable, given the cognitive heuristics and the selection procedure through which cases are examined (Chan 1979; Tversky and Kahneman 1974; Tversky and Kahneman 1973). I first present an argument about the cognitive biases that hinder retrospective intelligence examination, then argue for a more nuanced understanding of accuracy the requires examining times when no warning should have been issue. These two arguments inform the data I and empirical method I use later to test my

hypothesis.

Humans have well known and inescapable cognitive biases that make retrospective studies of intelligence very difficult (Tversky and Kahneman 1974; Kahneman 2011), even when people are made aware of their biases (Slovic and Fischhoff 1977).[1] Writing about intelligence, Chan (1979) points out the effect of the availability heuristic on how current observers rate past intelligence failures:

> Due to familiarity and salience, people more readily recall events that are recent, traumatic, and which happened to oneself or others close to oneself. With regard to the analysis of strategic warning, this bias leads to the over-representation of intelligence failures, threat assessment cases, and one's own experiences or those of one's allies in the recent past. Conversely, the result in the under-representation of intelligence success [...] particularly if they entail stretching the memory to the distant past (175).

Moreover, when people know the outcome of an event, they find it impossible to not let that knowledge cloud their view of it and are unaware of the degree to which their hindsight changes their view (Fischhoff 1975). These biases will cloud any retrospective look at past intelligence. Cases should be selected very carefully a confirmation bias of missed signals, and where possible, researchers should use objective, pre-specified measures of intelligence accuracy, applied transparently and consistently in evaluating past intelligence.

In the case of early warning, "accuracy" is best understood as consisting of two parts: recall and precision. Recall is a measure of how many of the cases that actually occurred had early warning before hand, and is the measure that's constructed through retrospective

[1] Almost all of the biases that Tversky and Kahneman find make it difficult to forecast in the first place, making this article immensely important reading for intelligence analysts. Among the analysis-limited biases they identify are people's complete discounting of base rates and prior probabilities, a "gambler's fallacy" that events that have not occurred in a long time are "due" to happen, an ignorance of what is predictable at all, the "availability" heuristic whereby easily conjured or remembered events are assessed to be much more likely, and excessive anchoring on initial values or beliefs (1974). Sadly, as fascinating as they are, a longer discussion of how these biases affect intelligence work is outside the scope of this paper. See also Kahneman (2011) and Tetlock (2005) for a discussion of cognitive biases and forecasting.

studies of particular events, e.g., the failure to warn in Pearl Harbor (Wohlstetter 1962), the success (perhaps) of early warning in Rwanda (Feil 1998), and the failure to adequately warn of the September 11, 2001 attacks (Kean 2011). Precision, on the other hand, measures how many warnings are correct. Failure to forecast an event (a recall problem) is very undesirable, but an overcompensation by crying wolf and decreasing precision also harms policymakers' ability to respond correctly (Betts 1980, Aesop c.560 BC). Bureaucratic politics, however, may pressure diplomats and intelligence agencies to over-warn, decreasing precision (which is undervalued) to increase recall (since failures to warn are extremely costly to them). Studying precision requires looking at a large set of predictions and assessing how many of them came to pass, a task that is impossible when picking specific cases that did come to pass and examining the warnings leading up to it.

Previous Research

This study fits into the early warning literature, defined broadly to include the literature on intelligence failures and the preventive diplomacy literature. I argue that this literature poses two fundamental questions: is conflict early warning possible? and why is there a gap between warning and response?

Is early warning possible?

Developing early warning technology for anticipating future conflict has been a staple of research for the defense and intelligence agencies, as well as academic researchers since the 1970s.[2] The literature on technical forecasting using structured data such as GDP, past conflict, and mountainous terrain, is relatively well developed in political science, but I put it aside

[2]In answering the question of how successful these techniques are, we only have access those that are publicly known, meaning it is possible that better, classified techniques could exist.

to focus on the smaller literature that finds early warning in text. A literature has emerged in political science over the last 15 years that examines whether the warning signs for future conflict are available in news text. It is important to note that this work simply uses text as a source of raw information to feed into new forecasting approaches. It does not look for actual warnings in the text itself, made by contemporaries. Advances in computing power and better access to large collections of digital text have made this work possible. The main findings of this literature are that news text contains useful information that can be used to predict future conflict.

The major approach here is the "event data" approach, which extracts descriptions of political events from text and uses them as inputs to forecasting models to try to predict civil war or other conflict, with some success (Schrodt 2006; Brandt, Freeman, and Schrodt 2011; D'Orazio, Yonamine, and Schrodt 2011 Shellman, Levey, and Young (2013); Garces, Lustick, and Levey 2015; Beger, Dorff, and Ward 2014). These models find factual descriptions of important political events in text and include them in quantitative models to develop their own measures of warning. Another, more recent approach uses the contents of the text more directly to estimate tensions or precursors to conflict (Chadefaux 2014). Calculating simple count of the number of synonyms for tension in news archives going back to 1902 and adding that to a baseline model for interstate war onset improved the accuracy of the model for a war beginning the next week from 36.9% to 61.4% (Chadefaux 2014, 12).[3] This research, along with the structural civil war forecasting literature, show that civil wars are predictable, and that the contents of unclassified reporting provide very valuable information to help predict it.

[3]Ongoing work by Benjamin Valentino, Jay Ulfelder, and Andrew Halterman is extending this measure to better predict onsets of new episodes of state-led mass killing (paper to be presented at ISA 2016).

Why is there a gap between warning and response?

Warning is possible, yet it is not always heeded. George and Holl (1997) argue that the early warning literature focuses excessively on the accuracy and timeliness of warning and ignores the problem of getting policymakers to decide to act on the information. They propose the term "warning-response gap" to emphasize that response is far from guaranteed even when early and accurate warning is provided. They emphasize the costs of policymakers of receiving early warning, specifically that they then have to do something about it:

> Thus, in many ethnic and religious conflicts, humanitarian crises, or severe human rights abuses, *timely or accurate warning may not be the problem at all.* [...] Moreover, this reaction can occur not only when what is at stake is only dimly perceived or not foreseen at all, but also if the coming crisis is *fully and accurately anticipated*. (George and Holl 1997, 9 (emphases in original))

Fifteen years later, Meyer et al. (2010) argue that the concept of the "warning-response gap" has not been developed as fully as it should, that decision-making, not technical improvements to forecasts, are the greatest obstacle to early action (557).

This argument in the 1990s literature on humanitarian early warning is predated, though, by research in the older surprise attack literature. Betts, writing in this literature, sees five stages in responses to surprise: data collection, the intelligence agency's decision to warn, the communication of the warning to decision makers, the decision to respond, and the (military) implementation of the response (1980, 552). Technical forecasting of civil wars and humanitarian emergencies addresses the first two states (Meyer et al. 2010), to the detriment of its ability to effect real responses to crises. Practitioners building early warning systems need to also address the strategic, cognitive, and normative concerns of decision makers in an attempt to persuade them to respond. From a research perspective, though, the "warning-response gap" is an empirical claim predicated on the existence of good, accurate, and timely early warning. Whether this occurs or not inside the government is an open question.

Data

To study the precision of early warning and to look for evidence of over-warning, I need a comprehensive set of potential warning documents. To provide this, I use the corpus of declassified State Department cables from 1973-1978, consisting of all cables sent inside the State Department, with the exception of a small number of Top Secret cables that remain unreleased. These cables, 1.8 million in total, were declassified and released between 2006 and 2014. I scraped, formatted, stored, and made searchable the 1.5 million that were complete (around 300,000 did not include text when given to the national archives or had unparseable date fields). See Appendix B for technical details.

This cables are often known as the "Kissinger cables", as he served as Secretary of State and National Security Advisor for almost the entirety of this period and was a prolific writer of cables himself, authoring over 140,000.

Table 1: Cables In Dataset Sent By Kissinger

Classification	Number of Cables
Unclassified	85,326
Limited Official Use	33,909
Confidential	24,309
Secret	46
Top Secret	0
Total	**143,590**

Some limited scholarly work on this corpus exists already. In computer science, Esteva et al. (2013) use the cable corpus to test advances related to automated document sorting,

attempting to infer the original security classification and geographic/subject tags on the cables using machine learning approaches. Zipple (2014) examines the U.S. response to and awareness of the anti-communist "Red Drum Murders" in Thailand in 1972 using the set of declassified cables pertaining to Thailand, finding that the State Department has limited awareness of the atrocities but was supportive of the Thai military despite evidence of their brutality. Some cables from this period were declassified earlier and used extensively in research on U.S. clandestine operations in Chile and on the U.S.-backed Pinochet coup.

Early warning in the cables

A full-text search of the cables gave me some confidence that they would cover the issues under study. The results of several searches are in Table 2.

Table 2: Early Warning Search Terms in the Cable Corpus

Search Phrase	Number of Cables
civil war	2,817
genocide	639
early warning	1,806
threat of civil war	15

On further examination, many of these cables were not relevant to this study. Many unclassified cables are simply transcripts or clippings of local speeches, broadcasts, and papers, and quote without comment people who allege genocide or "the threat of civil war". Most of the cables containing the term "early warning" that I examined were talking either about early warning of Soviet attacks in Europe or early warning indicators for economic changes.

One cable in particular gave me confidence that I would easily find many others. A cable sent on February 7, 1975 from the American consulate in Luanda predicted the onset of the Angolan civil war in the first line: "Although still isolated and not yet self-sustaining, recent outbreaks of physical and verbal conflict between MPLA and FNLA raise the spectre of an Angolan civil war."[4] It concludes with a frank discussion of U.S. interests at stake in Angola, primary of which is "the avoidance of a civil war which could bring requests for U.S. involvement" and threaten regional stability, U.S. commercial interests, and U.S. Navy refueling facilities. However, even this cable has some problems as an example of early warning. It was written 10 months before the onset of the Angolan civil war in December, and thus is of debatable accuracy. Furthermore, "raise the spectre" is a phrase that cannot be easily translated into an actual probability, which is well in line with the evasive language that Tetlock (2005) finds in his study of political prognosticators and Kent (1964) found in his work at the CIA.

The list of civil war onsets I use comes from the UCPD/PRIO Armed Conflict Dataset (Themnér and Wallensteen 2014; Gleditsch et al. 2002), which includes start dates for intrastate war that result in more than 1,000 battle deaths. I would have also liked to look at early warning for mass killing, to better test claims of foresight that come up around Rwanda, but I was unable to find a dataset of mass killing episodes that went back to the 1970s and included days or months of onsets, rather than years. The U.S. government-funded Political Instability Task Force provides a dataset of state-led mass killing episodes, of which 11 occur in this period. Their data is only coded to the country-year, however, which does not offer high enough resolution to examine early warning dynamics in the months immediately before the beginning of the mass killing episode.

[4]U.S. Consulate Luanda Cable 1975LUANDA00157 To U.S. State Department, "The Possibility Of Civil War In Angola During The Transitional Period And Policy Implications", 7 February 1975, SECRET, LIMDIS. (Declassified 05 JUL 2006)

Data Limitations

Using these cables to study civil war early warning creates two unavoidable problems:

First, these cables cover only a short historical periods—specifically, a period before the emergence of much of the contemporary concern with preventing or mitigating civil wars and atrocities. As a result, findings derived from this set of cables may not map well onto a post Cold War context with greater concern for and awareness of intrastate war. Early warning researchers in the 1970s were much more concerned with strategic warning, i.e., predicting imminent attack by the Soviet Union (Betts 1978; Chan 1979; Betts 1980). The unclassified sources of information that the U.S. government can now draw on for civil war warning have also greatly expanded: the U.S. government now has access to information from conflict zones coming from humanitarian NGOs, whose presence in conflict areas exploded after the 1990s (Stoddard 2006). A finding in this research that no warning or over-warning occurred in this period would not necessarily indicate that the same dynamics are at work in the 1990s and onward.

Second, the State Department is not primarily an intelligence-gathering organization and is perhaps the junior partner in the U.S. government's information gathering apparatus. A failure by the State Department to warn of impending war or mass killing does not by itself imply that the U.S. government did not have any advance warning. Any claims about what decision makers knew when based solely on an analysis of State Department cables will necessarily be limited. Claims about the State Department's patterns of under- or over-warning, in contrast will be well supported.

On the question of the State Department's role in civil war early warning, we have differing information from the period. Writing in his memoirs, Kissinger describes the CIA Director as being most important in early warning:

> It is the Director on whom the President relies to supply early warning; it is to the Director that the Assistant first turns to learn the facts in a crisis and for the analysis of events. And since decisions turn on the perception of the consequences of actions, the CIA assessment can almost amount to a policy recommendation. (1979, 487)

However, in a study at the same time about forecasting in the U.S. government, Phillips and Rimkunas (1979) find that it is the State Department that has primary concern for intra-state instability and civil war: the CIA during this period was more concerned with communist insurgencies against friendly governments and in limited the spread of communism. Knowing the exact distribution of capabilities and intelligence would require impossible access to the CIA's archives, though a thorough examination of National Security Council archives as they become declassified could shed light on this question.

Despite these time period and bureaucratic limitations, I argue that the cable corpus is worth studying for what it can tell us about early warning. As I argue above, assessments of intelligence need to examine periods where no conflict broke out and thus no warning was warranted, in order to make claims about the "batting average" of intelligence (Marrin 2012). As the only set of documents that offer a comprehensive coverage of all potential warning in a time period, the State Department cable corpus is unmatched.[5]

Questions and Observable Implications

Using this data, I seek to resolve the puzzle posed above by answering two questions:

1. Is there such a thing as a "warning cable", analogous to the stylized "genocide fax" in Rwanda? If cables providing an unambiguous early warning exist, they should have features of their text that are learnable and recognizable.

[5]The Wikileaks cable corpus is the only other available corpus that could be used for this purpose, but technically remains classified and out of bounds for me.

2. As a weaker test, is there any predictive power in cables? Do pre-civil war cables look different in some systematic way from those in other countries or at other times? If so, an automated text classifier should be able to find differences in in the words and phrases that are used in the different periods.

Method

To answer these two questions, I used automated text analysis techniques to try to identify consistent markers of "warning cables" and to find the different characteristics of cables from pre-war phases. Using automated techniques, rather than only working by hand, has two advantages. The first advantage is that automated techniques attenuate some of the psychological biases that I discuss in the literature review section. As someone who knows whether war occurred or not in the countries described by the cables, I am prone to letting that knowledge cloud my assessment of whether they got it "right" or not, even when I believe that I am not. If a computer can classify the cables as warning or not, that would be very strong evidence that some cables are indeed smoking guns. The second advantage of an automated approach is simply scale. Assuming I can read a cable in 30 seconds, and assuming a grad school reading schedule of 80 hours/week, it would take more than 100 years to read all of the cables in this corpus. Some form of machine assistance is necessary for any task that involves searching the corpus for a particular kind of cable. Both of the methods I employ use a "bag-of-words" approach that ignores the word order in the document and simply counts the usage of individual words in the cables and on "bi-grams"–two word combinations that occur often in the corpus, such as "civil war". While ignoring word order throws away some of the meaning in the text, it is a standard technique that still results in remarkably good performance (Joachims 2002) A technical description of this process can be found in Appendix B.

Looking for Warning Cables

To answer question 1, I sampled and read over 300 cables, selected because they included terms that would make them higher likelihood of being civil war warning cables, such as "civil war imminent" and "rebel group". I hand-coded these as warning cables or not warning cables, using the criteria that warning cables had to make a strong statement of probability ("likely", "probable") and had to warn of future violence. Of the cables, 248 were not warning cables, 5 were warning cables, and an additional 51 were "near misses", that warned of escalation of existing violence or used vague terms or conditionals ("if x continues, no indication that y will not also continue…".

Even the cables I classified as warning cables do not present unambiguous predictions of likely future war, and were only included because I relaxed somewhat my criterion that the cable had to make an explicit statement about likelihood. For instance, note the hedging language in this cable about the role of the military in Chile's government in the lead-up to the CIA-supported coup in June (my italics):

> TENSE POLITICAL AND ECONOMIC SITUATION HAS LED TO PERSISTENT REUMORS FEATURED IN DAILY PRESS THAT MILITARY WILL SOON TAKE MAJOR ROLE IN GOVT. ATMOSPHERE OF DEMONSTRATIONS AND VIOLENCE, SERIOUS EXECUTIVE/ CONGRESS AND EXECUTIVE/ JUDICIARY QUARRELS, REPORTED UNREST IN POLICE CORPS, EXPOSES OF CORRUPTION, AND MILITANT STATEMENTS BY UP ORGANIZATIONS ALL CHARACTERIZE CURRENT SCENE. […] *POTENTIAL FOR* MAJOR CONFRONTATION GROWING OUT OF CONFUSED AND AGITATED POLITICAL SCENE CLEARLY EXISTS. *HOWEVER*, CHILE HAS SHOWN IT CAN WITHSTAND A *GOOD DEAL* OF CONFUSION, AGITATION AND UNREST.[6]

Even if on its face this cable warns of potential future violence, it is written in such a way

[6] American Embassy Santiago Cable 1973santia02589 to U.S. State Department, "MILITARY BACK IN THE CABINET?", 14 June 1973, CONFIDENTIAL (Declassified 30 June 2005).

that it could be seen as prescient regardless of the actual outcome.[7]

Another cable does present a stark warning, but is actually a letter written by General Romeo Espino, Chief of Staff of the Armed Forces of the Philippines, and passed along with favorable comment by a foreign service officer. Quoting from Espino's letter contained in the cable:

> THERE ARE CLEAR INDICATIONS THAT THE PURPOSE OF THE REVOLUTIONARY GOVERNMENT IS TO ESTABLISH AN INDEPENDENT STATE IN MINDANAO AND SULU. […] THE ACTIVITIES OF THE MUSLIM REBELS IN MINDANAO DO NOT THREATEN THE STABILITY OF THE REPUBLIC. HOWEVER, THEY MAY MAKE POSSIBLE THE SEQUESTRATION OF PHILIPPINE TERRITORY AND THE EXDRCISE OF EXTRANEOUS SOVEREIGNTY OVER FILIPINOS AND OTHERS LIVING THEREIN.[8]

Other cables present "early" warning in the sense that they were sent before the UCDP/PRIO civil war onset date, but provide warning too late in a series of events to allow for any preventative action, in an illustration of "surprise despite warning" (Betts 1980):

> WHAT BEGAN AS AN APPARENTLY SPONTANEOUS UPRISING BY ELEMENTS OF THE ETHIOPIAN ARMY SECOND DIVISION AND SECOND AIR GROUP IN ASMARA SPREAD RAPIDLY TO OTHER MAJOR UNITS OF THE ARMED FORCES, INCLUDING THE NAVY HEADQUARTERS IN MASSAWA, THE AIR FORCE HEADQUARTERS AT DEBRA ZEIT, AND FINALLY THE ARMY HEADQUARTERS UNITS IN ADDIS ABABA. AS IT GREW IN SCOPE THE REBELLION DISPLAYED SIGNS THAT IT WAS WELL-ORGANIZED AND COORDINATED AND HAD PERHAPS BEEN PLANNED IN ADVANCE, AND THAT IT WAS AN OUTGROWTH OF A DISSIDENT MOVEMEN (sic) ORGANIZED DURING THE PAST TWO MONTHS BY JUNIOR OFFICERS.[9]

[7](Although the embassy could be accused of analytic "insider trading" here, given that it was involved in fomenting at least some of of the unrest. (Select Committee to Study Governmental Operations with Respect to Intelligence Activities 1975))

[8]American Embassy Manila Cable 1973MANILA02769 to U.S. State Department, "PHIL/MUSLIM CONFLICT: GOP MILITARY ASSESSMENT AND URGENT REQUEST FOR MAP AID," 9 March 1973, SECRET (Declassified 30 June 2005).

[9]U.S. State Department Cable 1974STATE042218, "REPORT ON ETHIOPIAN SITUATION," 2 March 1974, SECRET (Declassified 30 June 2005).

A cable describing renewed fighting again does not present predictions unalloyed with hedging language:

> A COMPLICATE (sic, throughout) SERIES OF MULTI FACETED NEGOTIATIONS IS NOW IN PROGRESS INVOLVING CHAD, LIBYA, NIGER, SUDAN AND THE VARIOUS REBEL GROUPS. PROSPECTS FOR SUCCESS IN THESE NEGOTIATIONS ARE OBSCURE AND HOSTILITES *MAY WELL* RESUME IN THE NEAR FUTURE. IF SO, DESPITE ADDITIONAL FRENCH MILITARY ASSISTANCE (WHICH MIGHT CHANGE AS A RESULT OF THIS MONTHS ELECTION), CHAD IS IN A WEAK POSITION TO DEFEND ITSELF. ACCORDINGLY, THE GOVERNMENT WITH WHICH WE MADE OUR COMMITMENT OF ASSISTANCE *MAY WELL* CEASE TO EXIST BEFORE WE HAVE HONORED IT.[10] (emphasis mine)

Although none of the cables besides the Angolan civil war cable present a strong warning of future civil war, I wanted to check whether they had word traits in common in a way that could be identified by a computer. If so, this would allow me to find other warning cables from other periods without having to randomly sample and hope to find warning cables. I trained a text classifier (Pedregosa et al. 2011; Joachims 1998, see Appendix B) on that labeled data. The classifier treats each document as a collection of words and two-word phrases ("bigrams"). It treats each document as a point in a space defined by the counts of their words, and tries to find the boundary in this space that best separates the warning phase documents and the not warning phase documents (see Figure 1).

[10] American Embassy Ndjamena Cable 1978NDJAMENA00952 to U.S. State Department, "US GOALS AND OBJECTIVES IN CHAD," 6 March 1978, SECRET (Declassified 20 March 2014). This cable also displays a remarkably different view of the Peace Corps' purpose from the modern conception, seeing it as a powerful tool for stabilizing countries experiencing civil wars: "WE HAVE ALREADY TAKEN ACTION TO EXPAND PEACE CORPS PROGRAMS, PARTICULARLY IN RURAL DEVELOPMENT AND AGRICULTURE. THE PEACE CORPS IS INCREASING ITS VOLUNTEER STRENGTH FROM A CURRENT LEVEL OF 80 TO 115 NEXT SUMMER. VIRTUALLY ALL OF THE NEW ARRIVALS WILL BE ENGAGED IN AGRICULTURE AND RURAL DEVELOPMENT. THESE FIELDS INVOLVED 11 PCV'S IN FY 77; 22 IN FY 78; AND WILL INVOLVE ABOUT 50 IN FY 79. IN ADDITION TO THESE INCREASES, THE COORDINATION OF PEACE CORPS AND AID'S EFFORTS IS BEING AUGMENTED, SPECIFICALLY IN THE FIELDS OF WELL DIGGING, REFORESTATION, AND AGRICULTURE PROJECTS." (American Embassy Ndjamena Cable 1978NDJAMENA00952 to U.S. State Department, "US GOALS AND OBJECTIVES IN CHAD," 6 March 1978, SECRET (Declassified 20 March 2014))

Figure 1: Drawing A Separating Line Between Warning Phase Cables (Red) and Non-Warning Cables (Black) By Term Frequency

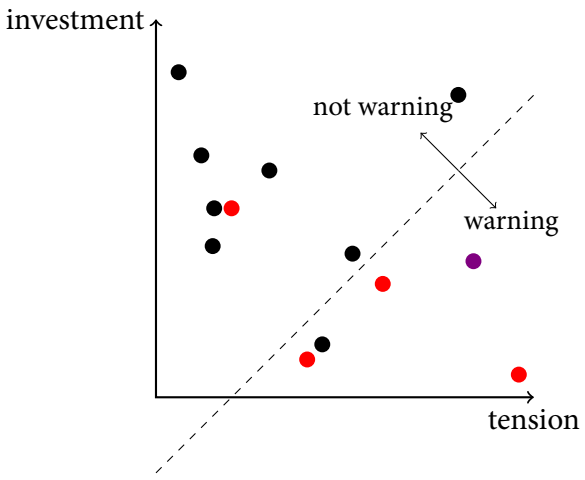

Figure 1 illustrates a very basic example of text classification. Assume that the only words we have to work with are "investment" and "tension". We can plot each document on a graph with one axis representing the number of times that "investment" is used in a document, and the other axis representing the uses of the word "tension" in a document. We can color each document according to whether it is a warning cable or not. The task then becomes to find the line that best separates the warning and not warning cables. Some documents will generally end up on the wrong side of the boundary for each word, meaning a degree of misclassification is inevitable. Previously unlabeled documents (purple) can be plotted and assigned a label based on which side of the line they fall on. In practice, this space has many more dimensions, one for each word, but the intuition remains the same.

After constructing the models and testing them on a held-out sample of labeled cables that had not been used in training the model, I can assess its accuracy on data it had not been trained on. Unfortunately, the ability of this classifier to match my hand-labeled cables was extremely poor. I then ran it on a large set of unlabeled data and examined the top 50 cables

by predicted probability of being warning cables. None of them were warning cables, so I concluded that my classifier did not work. It is possible that it would work better if I gave it more labeled data, but I believe that the issue is there there is no distinct "warningness" to pick up. The few warning cables I found were all different in their structure and the form in which they expressed their warning, making them difficult for a human to detect, let alone a computer. I was therefore unable to test an initial question about the distribution of warning cables, specifically, whether they occur more before actual civil war onsets, and whether they also occur before non-onsets, too. This would have helped to answer the "early warning" problem.

Looking for Signal in the Corpus

While I did not find evidence of a distinct characteristic of "warning cables", I still wanted to test whether cables contain any information that differentiates between the lead-up to civil war and times when civil war is not about to occur. I compared the $\sim$ 5,000 cables that were sent from countries where civil war broke out within 3 months with 5,000 cables from the same countries, but from the period encompassing 6 months before the civil war to 3 months before the civil war. This presents a very difficult test for the classifier. If I had given it non-civil war cables sampled at random from the entire corpus, it would have had the advantage of learning that descriptions of Western Europe are very good predictors of no civil war. Instead, the classifier has to learn the contents of the cables that change over from from the same country. Moreover, warning signs do not start neatly immediately before onset, meaning that warning may bleed over into the earlier phase and dilute our ability to find "warningness". Indeed, the paradigmatic Angola warning cable was sent well before the 3 month period I use here.

As an initial validation, I examined the words that were most useful in distinguishing these periods. If the classifier is actually picking up on "warningness", these should be

meaningful. I ran the classifier first without removing proper nouns, and it determined that the best predictor of civil war is where the following Foreign Service Officers were serving (Fisher, Shelton, Lavin, Killoran, Sober) and their countries (Bangladesh, Ceylon). After removing all proper nouns in the corpus, it found someone better terms, though not great: "debt," "liber[ation] movement," "fall bridg[e]," "captur[e]," "help," "fretelin," opportun[ity]," "struggl[e]," "troop," and "fight."^[Note that the proper noun remove is not perfect: "fretelin" remained in the cables despite being the name of the main Indonesian rebel group and is indeed highly predictive of civil war onset. Also note that the words have been "stemmed" to make them more easily counted. "captur" will subsume "captured," "capture," and "capturing,", all of which have essentially the same meaning in this context.

More quantitatively, comparing the 0-3 month vs. 3-6 month:

Table 3: Classifier for Whether Cable is 0-3 or 3-6 Months Before Civil War (Out of Sample Test)

	f1-score	# cables tested
not warning	0.64	1374
warning	0.62	1484
avg / total	0.63	2858

The F_1 score is a weighted average of two different accuracy measures. It tells us that a cable that the classifier labels as coming from a not warning phase is about 64% likely to have come from that phase, and that it correctly finds around 64% of the not warning cables. The same interpretation holds for warning phase cables. In absolute terms, this accuracy is poor, but given the difficulty of this task for the classifier (placing cables from the same country from around the same time on the correct side of an arbitrary 3 month cutoff), this perfor-

mance is remarkably good. Note that these cables are different from the ones used to train the classifier, so the algorithm has not simply "memorized" which cables belong where. Substantively, what this tells us is that despite being from the same country and likely written by the same foreign service officers, the cables coming from a country right before war breaks out are systematically different from cables from an earlier period, and this difference is detectable by a relatively simple machine learning algorithm.

Since this classifier was built to distinguish between the phase in a country that experienced a civil war right before the war from an earlier phase in that country, it may be able to find "early warning" signals in other countries. I ran the classifier over all the cables in the corpus that were not from a country that experienced civil war at some point. I then grouped the cables per sender for each month and took the average score from the warning classifier for all cables in that sender-month.

Table 4 shows the results. In somewhat of a vindication of for the "early warning" scholars of the 1970s, who studied surprise attack rather than humanitarian crises, the most common sender for these cables was the U.S. negotiators at the Mutual and Balanced Force Reduction talks in Vienna, which began in 1973 and concerned reductions in U.S. and Soviet conventional and nuclear forces in Europe. The other countries included here are general unstable or have experienced civil wars, with the exception of Belize, which has been almost entirely peaceful.

Conclusions

To a certain extent, this paper is a null finding. My initial hypotheses upon starting the research was that the State Department over-warned of coming civil wars, and that the poor precision of State Department warning debased the currency of warning and that this was

Table 4: Country Months With Highest Levels of Warning (Civil War Countries Excluded)

Month	Sender	Average "Warningness"	Cables Sent
1977-11-01	MBFR	0.61	36
1974-11-01	MBFR	0.60	59
1977-12-01	MBFR	0.60	26
1975-10-01	Swaziland	0.60	26
1974-10-01	MBFR	0.59	57
1976-10-01	MBFR	0.59	30
1975-11-01	Somalia	0.59	24
1974-10-01	Guyana	0.59	46
1975-10-01	Fiji	0.59	26
1975-10-01	MBFR	0.59	36
1975-10-01	Mozambique	0.59	49
1976-11-01	MBFR	0.58	24
1974-10-01	SALT II	0.58	37
1975-10-01	Belize	0.58	23
1975-11-01	Mauritius	0.58	21
1976-10-01	Lebanon	0.58	96
1978-11-01	MBFR	0.58	53
1977-11-01	Mauritius	0.58	27
1974-12-01	MBFR	0.58	42
1975-11-01	Botswana	0.58	34

in part responsible in part for a lack of preventative action. The failure of my cable classifier to accurately categorize "warning" cables precluded a test of this hypothesis. The reason for its inaccuracy could be a failure of the technology or a dearth of hand-tagged cables to learn from, but I believe it is because "warning cables" (even in a loosened definition) do not have a distinctive signature. Despite this null finding for my first question, the answer to my second question allows me to draw three conclusions, about the State Department's warning, about warning in general, and about warning for civil war and mass killing onset.

1. What does this tell us about the State Department warning?

The success in distinguishing between cables from the imminent war period from the period of 3-6 months before war indicates that civil wars are predictable to a degree, and that

foreign service officers are collecting and transmitting to Washington information that has predictive power. However, the utter inability to find consistent markers of a "warning cable", or even to find a warning cable according to the strict definition that it provides a strong statement of likelihood about large-scale future violence, seems to show that the State Department is not sounding the alarm in an obvious way. In part, this confirms what we already know, that the State Department is not primarily a "warning" or intelligence organization. This remains a key limitation of the study, that I can only study the State Department and not the intelligence community. In any case, the signal from the State Department about which places are about to have civil wars is weak. In the CIA, it may be consistent over-warning, but here it is consistent under-warning.

2. What does this tell us about warning in general?

The lack of clearly identifiable warning cables agrees with the argument I lay out at the beginning, that the retrospective hunt for "warning" is ad-hoc and suffers from cognitive obstacles. Without knowledge about how things turned out, there are no consistent warning signs. Or, as Kissinger himself put it,

> After every crisis there surfaces in the press some obscure intelligence report or analyst purporting to have predicted it, only to have been foolishly ignored by the policymakers. What these claims omit to mention is that when warnings become too routine they lose all significance; when reports are not called specifically to the attention of the top leadership they are lost in bureaucratic background noise, particularly since for every admonitory report one can probably find also its opposite in the files. (Kissinger 1979, 38).

Relatedly, from the 300 cables I hand-coded, it became clear that even the cables that seemed to warn about impending civil wars actually use hedging language to the point where they would seem correct regardless of war broke out or not ("HOSTILITES MAY WELL RESUME"). Warning may exist in isolated places, but finding it is a matter of interpretation and is clouded by cognitive heuristics and the knowledge of what happened next.

3. What does this tell us about intervention in civil wars and atrocities?

The reasons of intervention in an ongoing humanitarian crisis are multifaceted and well researched. They involve varying degrees of strategic calculations, norms and identities, domestic political concerns, and cognitive biases. When the war or crisis is only in its earliest stages, some of these factors become less important, and the issue of intelligence or early warning becomes paramount.

Other agencies may be able to predict civil wars better, but several pieces of evidence indicates that the classified realm is not much more advanced than the public state of the art in civil war forecasting. The first piece of evidence is that the Defense Department, specifically though the Defense Advanced Research Projects Agency (DARPA, or at times ARPA) has engaged in a series of publicly known efforts to improve forecasting accuracy since the 1970s. Early efforts relied too much on expensive human labor and did not produce decisively better forecasts (Singer and Wallace 1979; Schrodt 2012). A newer DARPA-funded project, the Integrated Conflict Early Warning System (ICEWS) [O'Brien (2010);icews2015], has better accuracy in predicting civil wars and has entered operational use, indicating that it has at least a modicum of utility over pre-existing methods.

On the intelligence community side, the long-running Political Instability Task Force (previously State Failure Task Force) has dramatically improved public forecasting methods for predicting civil war onset (Goldstone et al. 2010; Ulfelder 2012; Ward et al. 2013). The Central Intelligence Agency has continued to fund PITF, though whether the models built by PITF are in operational use is unknown. The Intelligence Advanced Research Projects Activity (IARPA), the intelligence community's version of DARPA, has very recently undertaken several large efforts to forecast societal instability, as a possible precursor to civil war. These include the Open Source Indicators program,[11] which analyzes social media and other

[11]http://www.iarpa.gov/index.php/research-programs/osi

text to predict protests, the ACE/Good Judgement Project, led by Phil Tetlock, which used human forecasting teams to greatly outperform benchmarks for forecasts of events including civil wars, and the recently launched Mercury program[12], which seeks to forecast political crises and civil unrest using automated processing of SIGNIT. These investments all seem to indicate that the IC and Defense Departments do not have a good sense of when civil war will break out either, making the question of whether the State Department could predict it highly relevant.

State Department cables include information that could be used to identify periods of greater likelihood for civil wars, but cables themselves do not provide good early warning. Factors like strategic interests, domestic political dynamics, and cognitive issues may explain (lack of) intervention in ongoing civil wars, but I think the puzzle of why so little early intervention when the benefits would be so great is best explained by poor early warning. Without early warning, the puzzled posed at the beginning of the paper becomes less confusing: early action to prevent or mitigate civil wars and atrocities does not happen in part because the warning never arrives.

[12]http://www.iarpa.gov/index.php/research-programs/mercury

Appendix A: Civil War Onsets

UCDP/PRIO Civil War Onsets, 1973-1978

1. Government of Bangladesh vs. JSS/SB, 1975-02-28
2. Government of Pakistan vs. BLF, 1973-05-18
3. Government of Angola vs. FNLA, UNITA, 1975-11-11
4. Government of Indonesia vs. Fretilin/East Timor, 1975-12-07
5. Government of Morocco vs. POLISARIO/Western Sahara, 1975-09-01
6. Government of Afghanistan vs. Jam'iyyat-i Islami-yi Afghanistan/Harakat-i Inqilab-i Islami-yi Afghanistan/PDPA 1975-12-31
7. Government of Nicaragua vs. FSLN, 1974-12-27
8. Government of Sri Lanka vs. LTTE, TELO, 1975-07-27
9. Government of Ethiopia vs. ALF, 1975-06-30
10. Government of Ethiopia vs. OLF, 1974-08-31
11. Government of Mauritania vs. POLISARIO, 1975-12-10
12. Government of Myanmar (Burma) vs. LNUP, 1973-01-11

Appendix B: Technical Details

The first step in the project was downloading the 1.8 million cables. All of the cables are released and available for download, but the only mechanism provided on the National Archives site for viewing them is a by-year keyword search option, that displays 10 results at once. I reverse-engineered the URLs system for each year, allowing me to iterate through the entire list of available cables for that year and download them. The cables are only available in PDF form.

The Python program to download each year is `cable_scrape.py`. For usage instructions, run:

```
python cable_scrape.py -h
```

This program will write all of the PDFs for that year to a directory and write diagnostic information to a log file. Because of the large number of files, this process can be very slow to run because of limitations on the speed of writing files to a hard drive. If possible, use a solid state drive (SSD), which is many times faster than a conventional spinning disk hard drive.

Code for processing the cables

After downloading, the cables are formated as individual PDF files with all the information contained in the next but in a semi-structured way. First, to convert the cables from PDF to .txt, I use a utility called `pdftotext`. On a Linux or Mac computer (assumed for this Appendix), the command to convert the all of the cables in a given directory from PDF to text is:

```
find -type f -name '*.pdf' | xargs --max-procs=4 -n1 pdftotext -layout
```

Using `ls` won't work because the file list is too long. That's what the pipe to xargs does. The `--max-procs` thing is the easy bash way of parallelizing it. The `-layout` flag tells `pdftotext` to try to preserve the formatting. This is important to keep things like the subject and sender from getting collapsed onto one line, making it very difficult to disentangle them.

Although the cables are not that large in aggregate (6 GBs), they are very unwieldy to work with because of their number. A database helps to get us around the problem of dealing with hundreds of thousands of cables, especially when they need to be searched. I use a NoSQL document datastore called Mongo to store the cables. The Mongo database should be started (after it is installed), with a special switch to enable full text searching:

```
mongod --dbpath /media/data/ --setParameter textSearchEnabled=true
```

The `cables_to_mongo.py` script can be used to load a directory of cables into the Mongo database. For instructions, run

```
python cables_to_mongo.py -h
```

After the cables are loaded in (this process should take much less than an hour if the cables and MongoDB are both on the same solid state drive), I constructed a full-text index to allow for quick searches of the text:

```
db.cables.createIndex({
                        text : "text",
                        sender : "text",
                        subject : "text"
                        }
                    )
```

Code to perform the full text searches used in Table 2.

```
db.cables.count({ $text: { $search: "\"civil war\"" } })
# 2817
db.cables.count({ $text: { $search: "\"genocide\"" } })
# 639
db.cables.count({ $text: { $search: "\"early warning\"" } })
# 1806 (mostly economic indicators and nuclear)
db.cables.count({ $text: { $search: "\"threat of civil war\"" } })
# 15 (mostly quotes of speeches by others)
```

Calculate the most common senders like this (from Python):

```
from pymongo import MongoClient
connection = MongoClient()
db = connection.text
collection = db["cables"]

senders = collection.aggregate([
                    { "$group":
                        {"_id":"$sender" ,
                        "number":{"$sum":1}}
                    }
                ])
sender_list = [i for i in senders]
```

```
newlist = sorted(sender_list, key=lambda k: k['number'], reverse = True)
newlist_df = pd.DataFrame(newlist)
newlist_df.to_csv("sender_list.csv")
```

The sender fields are partially inconsistent and have misspellings and alternative embassy/consulate names throughout. I created a mapping by hand of the top senders to the correct country. The mapping is available in `sender_list_labled.csv`.

The approach to answering Question 1 with hand tagged cables was a dead end. Instead, I answer Question 2 by looking at how the words used before civil wars differ from the words used regularly. I found all cables from a each country that experience civil war, in the period from 3 months to 1 day before the UCDP/PRIO onset date. I also found all cables from 6 months to 3 months before to use as a comparison. See example code below, or `warning_phase_from_mongo.py` for full code.

```
db.text2.count( {$and :
                    [{"cleaned_sender": "BANGLADESH"},
              {"iso_date" :{$gte : ISODate("1974-12-28-08T00:00:00Z")}},
              {"iso_date" : {$lte : ISODate("1975-02-28-08T00:00:00Z")}}
                    ]
                }
            )
```

As I discussed in the methods section on page 22, my first iteration of the algorithm that looks for differences in word usages between pre-civil war periods and other periods quickly picked up on the names of the countries where civil wars occurred and the foreign service officers who were serving in them. These of course are not real warnings of civil war, so I needed to exclude them. I re-wrote the `cables_to_mongo.py` to run a named entity recognition (NER) system over the cables to find any proper nouns for people, locations, or adjectives. The system I used was Lincoln Lab's MIT Information Extractor (MITIE)[13], which provides accurate and very fast named entity recognition in English. Unlike the text that MITIE is built to work with, the cables are all capitalized, which causes lower accuracy. I got better results by making all the text lower case, running MITIE over that, and re-capitalizing the nouns for storage.

After the named entities are stored in the database, they can be retrieved like this:

```
mongoexport --db text --collection text2 --csv --out all_entities.csv \
    --fields 'entities'
```

These names will be used later to remove all proper names from the dataset before training the model, to avoid just learning the names of foreign service officers in these countries as the best predictors of civil war.

[13] https://github.com/mit-nlp/MITIE

Text Analyis

These are the steps used to build an automated text classifier that distinguishes between different known categories of text (warning/not warning cables or pre-civil war cables vs. other time period cables). All of these tasks were done in Python with the scikit-learn module (Pedregosa et al. 2011). See the `Cable_Classifier.ipynb` file in the replication materials.

1. Remove all proper nouns from the text, using the comprehensive list of named entities generated above.

2. Fit a word "vectorizer" that matches each word to a column in a matrix. This step also removes all proper names, stems the words (e.g., "anger", "angers", "angrily" go to "anger"), and removes very common English words (see `model_pickler.py` for the code for this step and the following).

3. Fit a term frequency-inverse document frequency model to the cables, which weights each word for each cable by the number of times it appears in the cable, divided by the number of documents in the corpus that contain that word at least once. This upweights rare words or relatively common words if they are use many times, and helps make short and long cables comparable.

4. Fit several models to the tf-idf transformed, labeled cables. These models include a linear SVM (Joachims 2002), a stochastic gradient descent model with a modified Huber loss (similar to a support vector machine), and an elastic net model (a combination of lasso and ridge regression). The SGD model produced the best accuracy and returned features with the best face validity.

5. Test the classifier against a set of held-out labeled cables, that is, cables for which I know the right answer, but the classifier has not see yet. This allows us to measure whether the classifier has actually found a deeper or consistent measure of the phenomena we are investigating, as opposed to simply memorizing words that predict each cable well.

The Python code to evaluate the time period classifier is here:

```
X_train, X_test, y_train, y_test = cross_validation.train_test_split(
    both_transformed, both.label, test_size=0.4, random_state=0)
clf = SGDClassifier(loss='modified_huber', penalty='l2',alpha=1e-3,
                    n_iter=5, random_state=42).fit(X_train, y_train)
clf.score(X_test, y_test)
print(metrics.classification_report(y_test, clf.predict(X_test)))
```

References

Beger, Andreas, Cassy L Dorff, and Michael D Ward. 2014. "Ensemble Forecasting of Irregular Leadership Change." *Research & Politics* 1 (3).

Betts, Richard K. 1978. "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable." *World Politics* 31 (01): 61–89.

———. 1980. "Surprise Despite Warning: Why Sudden Attacks Succeed." *Political Science Quarterly*, 551–72.

Boutros-Ghali, Boutros. 1992. *An Agenda for Peace: Preventive Diplomacy, Peacemaking, and Peace-Keeping: Report of the Secretary-General Pursuant to the Statement Adopted by the Summit Meeting of the Security Council on 31 January 1992*. Vol. 1247. UN.

Brandt, Patrick T., John Freeman, and Philip A. Schrodt. 2011. "Real Time, Time Series Forecasting of Inter-and Intra-State Political Conflict." *Conflict Management and Peace Science* 28 (1): 41–64.

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

Chan, Steve. 1979. "The Intelligence of Stupidity: Understanding Failures in Strategic Warning." *American Political Science Review* 73 (01): 171–80.

D'Orazio, Vito, James E Yonamine, and Philip A. Schrodt. 2011. "Predicting Intra-State Conflict Onset: An Event Data Approach Using Euclidean and Levenshtein Distance Measures." In *69th Annual Meeting of the Midwest Political Science Association, Chicago, Il, USA*.

Esteva, Maria, Jeffrey Felix Tang, Weijia Xu, and Karthik Anantha Padmanabhan. 2013. "Data Mining for 'Big Archives' Analysis: A Case Study." *Proceedings of the American Society for Information Science and Technology* 50 (1): 1–10.

Feil, Scott R. 1998. "Preventing Genocide: How the Early Use of Force Might Have Succeeded in Rwanda." *Leadership and Conflict Resolution*, 304–88.

Fischhoff, Baruch. 1975. "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty." *Journal of Experimental Psychology: Human Perception and Performance* 1 (3). American Psychological Association: 288.

Garces, Miguel, Ian S Lustick, and Brian P Levey. 2015. "Fragility and Destabilizing Protest: Combining Event and Structural Data for Improved Forecasts." *Procedia Manufacturing* 3: 4136–43.

George, Alexander L, and Jane E Holl. 1997. *The Warning-Response Problem and Missed Opportunities in Preventive Diplomacy*. Carnegie Commission on Preventing Deadly Conflict.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research* 39 (5): 615–37.

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

Gourevitch, Philip. 1998. "The Genocide Fax." *The New Yorker*.

Jentleson, Bruce W. 2000. "Preventive Diplomacy: A Conceptual and Analytic Framework." *Opportunities Missed, Opportunities Seized: Preventive Diplomacy in the Post-Cold War World*. Rowman & Littlefield Publishers, Inc. Lanham, 3–20.

Jervis, Robert. 1968. "Hypotheses on Misperception." *World Politics* 20 (03): 454–79.

Joachims, Thorsten. 1998. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. Springer.

———. 2002. *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Macmillan.

Kean, Thomas. 2011. *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States*. Government Printing Office.

Kent, Sherman. 1964. "Words of Estimative Probability." *Studies in Intelligence* 8 (4): 49–65.

Kissinger, Henry. 1979. *White House Years*. Little, Brown,; Co.

Kuperman, Alan J. 2004. *The Limits of Humanitarian Intervention: Genocide in Rwanda*. Brookings Institution Press.

———. 2013. "A Model Humanitarian Intervention? Reassessing NATO's Libya Campaign." *International Security* 38 (1). MIT Press: 105–36.

Lyons, Terrence, and Ahmed Ismail Samatar. 1995. *Somalia: State Collapse, Multilateral Intervention, and Strategies for Political Reconstruction*. Brookings Institution Press.

Marrin, Stephen. 2012. "Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure?" *Intelligence and National Security* 27 (6): 896–912.

Meyer, Christoph O, Florian Otto, John Brante, and Chiara De Franco. 2010. "Recasting the Warning-Response Problem: Persuasion and Preventive Policy." *International Studies Review* 12 (4): 556–78.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12 (1): 87–104.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

Phillips, Warren R., and Richard V. Rimkunas. 1979. "A Cross-Agency Comparision of U.S. Crisis Perception." In *To Augur Well: Early Warning Indicators in World Politics*, edited by J. David Singer and Michael D. Wallace. Sage Publications, Inc.

Posen, Barry R. 2014. *Restraint: A New Foundation for US Grand Strategy*. Cornell University Press.

Power, Samantha. 2013. *A Problem from Hell: America and the Age of Genocide*. Perseus Books Group.

Sambanis, Nicholas. 2002. "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War." *Defence and Peace Economics* 13 (3): 215–43.

Schrodt, Philip A. 2006. "Forecasting Conflict in the Balkans Using Hidden Markov Models." In *Programming for Peace*, 161–84. Springer.

———. 2012. "Precedents, Progress, and Prospects in Political Event Data." *International Interactions* 38 (4): 546–69.

Select Committee to Study Governmental Operations with Respect to Intelligence Activities. 1975. "Covert Action in Chile, 1963-1973. Staff Report of the Select Committee to Study Governmental Operations with Respect to Intelligence Activities." In *United States Senate*.

Shellman, Stephen M, Brian P Levey, and Joseph K Young. 2013. "Shifting Sands: Explaining and Predicting Phase Shifts by Dissident Organizations." *Journal of Peace Research* 50 (3): 319–36.

Singer, J. David, and Michael D. (eds.) Wallace, eds. 1979. *To Augur Well: Early Warning Indicators in World Politics*. Sage Publications, Inc.

Slovic, Paul, and Baruch Fischhoff. 1977. "On the Psychology of Experimental Surprises." *Journal of Experimental Psychology: Human Perception and Performance* 3 (4): 544.

Stanton, Gregory H. 2012. "The Rwandan Genocide: Why Early Warning Failed." *Journal of African Conflicts and Peace Studies* 1 (2): 3.

Stedman, Stephen John. 1995. "Alchemy for a New World Order: Overselling 'Preventive Diplomacy'." *Foreign Affairs* 74 (3): 14–20.

Stoddard, Abby. 2006. *Humanitarian Alert: NGO Information and Its Impact on US Foreign Policy*. Kumarian Press Bloomfield, CT.

Tetlock, Philip. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.

Themnér, Lotta, and Peter Wallensteen. 2014. "Armed Conflicts, 1946–2013." *Journal of Peace Research* 51 (4): 541–54.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207–32.

———. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

Ulfelder, Jay. 2012. "Forecasting Political Instability: Results from a Tournament of Methods." *Available at SSRN 2156234*.

Ward, Michael D, Nils W Metternich, Cassy L Dorff, Max Gallop, Florian M Hollenbach, Anna Schultz, and Simon Weschle. 2013. "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction." *International Studies Review* 15 (4): 473–90.

Western, Jon. 2005. *Selling Intervention and War: The Presidency, the Media, and the American Public*. Johns Hopkins University Press.

Wohlstetter, Roberta. 1962. *Pearl Harbor: Warning and Decision*. Stanford University Press.

Zipple, Matthew. 2014. "Thailand's Red Drum Murders Through an Analysis of Declassified Documents." *(SERAS) Southeast Review of Asian Studies* 36: 91–111.