Early Warning Signals for War in the News

Thomas Chadefaux

ETH Zurich
Department of Humanities, Social and Political Sciences
Clausiusstrasse 50, CLU C6
8092 Zurich, Switzerland
Phone: +41 44 632 04 58
fax: +41 44 632 17 67
thomasc@ethz.ch

July 19, 2012

Abstract

There have been more than 200 wars since the start of the 20$^{\text{th}}$ century, leading to about 35 million battle deaths. However, efforts at forecasting conflicts have so far performed poorly for lack of fine-grained and comprehensive measures of geopolitical tensions. Here, we developed a weekly risk-index by analyzing a comprehensive dataset of historical newspaper articles for 166 countries over the past century, which we then tested on a data of all conflicts within and between countries recorded since 1900. Using only information available at the time, we could predict the onset of a war within the next year with up to 85% confidence; we also forecasted over 70% of large-scale wars, while issuing false alarms in only 16% of observations. Predictions were improved up to one year prior to interstate wars, and six months prior to civil wars, giving policy-makers significant additional warning time.

> *There will be no European general war. [...] The six great powers—Germany, Austria and Italy on one side, and Russia, France and Britain on the other side—cannot afford a clash of arms.[...]. [They] will hesitate at the last moment and endeavor to adjust matters.*
>
> — Los Angeles Times, July 27, 1914

1 Introduction

Up until the very outbreak of World War I on July 28, 1914, newspapers took little notice of the rising tensions and the brewing conflict in Europe. In fact, in the week preceding its onset, worldwide newspapers mentioned "tensions" or "conflict" no more than at almost any time during the previous fifteen years. In other words, WWI seems to have come largely as a general surprise. This is in sharp contrast with World War II, for which the rise of tensions was echoed by a steadily growing attention from the press since at least 1935.

These differences in the anticipation of war are striking and raise a number of questions. Do wars usually come unexpected, or is the buildup of tensions visible and the outbreak of conflict predictable? Are there systematic differences in our ability to anticipate wars—for example, are conflicts with high levels of casualties easier to anticipate than the relatively costless ones? Or perhaps the type of war—inter or intra-state—is the determining factor? And could we have, using only information available at the time, derived earlier warning signals for war?

Unfortunately, the prediction of war has been the subject of surprisingly little interest in the literature, in marked difference to a wide range of fields, from finance to geology, which devote much of their attention to the prediction of extraordinary—"black swan" (Taleb 2011)—events such as financial crises or earthquakes. A recurrent difficulty has been the absence of a measure of tensions that is both fine-grained and comprehensive (Holsti 1963, Newcombe, Newcombe & Landrus 1974, Choucri 1974). Historical studies of single wars abound, but they are hardly quantifiable, rely on hindsight, or ignore the equally important cases in which war did not occur (Leetaru 2011). Others have focused on the conditions that are most conducive to war, but the indicators used are typically yearly, thereby missing the escalation of tensions and the timing of the conflict outbreak (Beck, King & Zeng 2004, De Marchi, Gelpi & Grynaviski 2004, Beck, King & Zeng 2000, Gleditsch & Ward 2011). In addition, these indicators are often poorly harmonized across countries, and their estimation (e.g., military spending) often depends on the government's goodwill or strategic interests. Finally, they cannot reliably measure the *perceived* reality of the time. Contemporaries may have been oblivious to real risks factors or, on the contrary, might have imagined them where none existed (Holsti 1963).

Here, we hope to fill this gap by deriving, for a large number of countries and times, a comprehensive estimate of tensions—a situation of stress and latent hostility—within and between countries, and of their perceptions by contemporaries. We do this by analyzing a large data of historical newspaper articles. The press is an ideal source of information because it provides fast, accurate and in-depth coverage of rising tensions throughout the world, and has a strong incentive to report any increase in the likelihood of war—whether perceived or real—since "what sells a newspaper" is war.[1] Moreover, news reports are written by journalists whose reputation is based upon the provision of accurate information. A database of news also avoids the problem of hindsight by using only information that was available at the time and, by consistently applying the same methodology to every war, avoids any temptation to cherry-pick the evidence. Finally, newspapers have an important advantage over event-base data: they can report tensions even when no actual event occurred (and hence nothing is recorded in the MID or COPDAB data). Conversely, an event might occur but not be perceived as significant by its contemporaries. In other words, an analysis of news gives us information about the interpretation of events by their contemporaries, and not a simple event description from which meaning needs to be inferred *a posteriori*, with the benefit of hindsight.

Of course, expert political opinions such as those of journalists might not be reliable sources of predictions (Tetlock 2005). However, we do not look for journalists predictions per se, but rather for indications of rising tensions in the aggregation of news coverage. Moreover, the aggregation of these reports and opinions takes advantage of the wisdom of crowds, and can hence provides more valuable information than any single expert opinion

[1] Lasswell (1971, p.192), cited in Ferguson (1999, p.11)

(Surowiecki, Silverman et al. 2007).

The resulting data is a fine-grained and direct proxy for the evolution of tensions in each country. We use it to derive an estimate of the probability of a coming war, which we then test on existing conflicts datasets—including all inter-, intra- and extra-state conflicts recorded with a starting date of January 1902 to December 2010.

We ask four main questions in this paper. First, do newspaper articles report growing tensions, or does war usually come as a surprise? And do the estimates of tensions correlate with the proximity to war, or are there generally no early warning signals for war in the news? Second, can we derive a reliable risk-index from an analysis of newspapers (hence using only information available at the time)? And would this index improve the predictions we would have made using only yearly variables such as military spending, or regime type? Third, are different types of war better predicted than others? For example, are interstate conflicts easier to forecast than intrastate ones? Or perhaps large wars can be anticipated earlier than small ones? Finally, we ask how far ahead news can provide warning signals. In other words, how far ahead does the analysis of news improve our predictions over a simple analysis based on yearly variables?

This paper proceeds in 4 steps. We first review the relevant literature on predicting conflicts. We then present a new data set on tensions, collected by analyzing a large database of newspapers. We also explain the conflict data on which the tensions-estimates will be tested. In section 4, we show that, typically, the number of reports about tensions significantly increases well ahead of a conflict, and that the number of conflict-related news is a significant predictor of conflict. Finally, we show in section 5 that early warning signals can be derived from this data and used as reliable predictors of wars, using only information available at the time. We also analyze the type of war best predicted—by type and casualties—and how far ahead news count can provide information.

2 Related Literature

The vast majority of research on forecasting war consists of a myriad of historical studies of single wars.[2] These accounts are invaluable for their depth of information and level of analysis: they identify the relevant actors, the in-depth political and economic issues, and rely on a wide variety of sources, such as newspapers, diaries, or international agreements. They also provide a thorough understanding of the international and local context, the central actors' personalities, and the institutional setting in which decisions were made. However, historical accounts suffer from intrinsic limitations that hinder systematic inferences. First, the process of historical analysis is highly time-consuming, so that deriving consistent conclusions about a large number of wars is an impossible undertaking by a single scholar. A solution would be to aggregate the accounts of many historians, but their

[2]See, for example, Ferguson (1999).

methods and emphasis are usually very different, and their results hardly quantifiable, and hence are difficult to compare. Moreover, historians work with the advantage of hindsight. As a result, they tend to focus on cases in which a conflict *did* occur, ignoring those in which rising tensions did not lead to war—the dog that did not bark. Thus, for the hundreds of books about World War I, only few focus instead on the absence of outbreak in 1913. Moreover, hindsight allows scholars to look *a posteriori* for specific evidence of mounting tensions, even if none stood out at the time.

In contrast, the international and comparative conflict literature has taken a more systematic approach by deriving the *conditions* most conducive to war. Arms races (Glaser 2000), long-standing territorial rivalries (Huth 1998), large and rapid shifts (Powell 2004) in power or rough terrain (Fearon & Laitin 2003) are some of the factors that have been associated with an elevated risk of conflict, either internationally or domestically. This approach has the advantage of identifying the root causes of tensions, and we do incorporate some of these variables in our models. However, the indicators used are typically yearly, thereby missing important parts of the escalation and the timing of the conflict outbreak. A related approach to the measurement of tensions relies on proxies—variables that change as tensions ebb and flow. For example, arms races or escalation steps (e.g., speeches, troop mobilization) provide important indicators of rising antagonism (Newcombe, Newcombe & Landrus 1974). However, they also suffer from drawbacks that can make them impractical. For one, these variables are often limited to interstate wars. Arms races, for example, are difficult to detect in non-state actors. Moreover, they are rarely harmonized across countries and might be imprecise (Lebovic 1998, Lebovic 1999), in addition to being usually only published on a yearly basis, rendering impossible more fine-grained analyses of the rise of tensions in the weeks and months preceding war. The estimation of these variables also often depends on the actors' goodwill or, worse, their strategic interests. Arms spending figures, for example, are typically released by the central government, and can be tweaked one way or another to serve a particular purpose—to intimidate or to reassure, for example.

Others have attempted to quantify tensions more directly. Thus the Conflict and Peace Data Bank is a "library of daily international and domestic events or interactions" (Azar 1980) and the World Events Interaction Survey "a record of the flow of action and response between countries (as well as non-governmental actors, e.g., NATO) reflected in public events reported daily in the New York Times from January 1966 through December 1978" (McClelland 1984). While these data document relevant events such as international border clashes or domestic press censorship with sufficient frequency and detail, coding is labor-intensive and the data's time-coverages are, as a result, limited (see also Weidmann & Ward (2010)) .[3] Other data do cover a longer time span, but at the cost of lower precision, recording the actual manifestations of conflict such as crises (Leng 1987), militarized interstate disputes (MID) (Gochman & Maoz 1984), and international wars (Sarkees &

[3] 1948–78 for COPDAB, and 1966–1978 for WEIS.

Wayman 2010*a*), rather than documenting continuously emerging tensions.

While these problems have long been recognized (Holsti 1963, Köhler 1975, Newcombe, Newcombe & Landrus 1974, Choucri 1974), there still exists no fine-grained, comprehensive data of tensions yet. We believe that an aggregate analysis of news articles can serve such a function. This idea is not entirely new. Karl Deutsch already recognized the importance of the mass media in mobilizing public opinion, and argued that a careful analysis of the media could yield early warning signals for interstate conflicts (Deutsch 1957, George 1956, George 1959). This insight was exploited systematically in Hunt (1997), which provides a methodology for identifying a regime's intention to launch a conflict in advance of the actual initiation using media analysis. Hunt identifies editorials in prominent newspapers closely tied to the regime as predictive indexes of the potential to go to war against a specific enemy. This approach is in a way more refined than ours here, since he uses information about the content of the article (in particular, was the tone toward another country critical?). However, this strategy also involves significant human coding and, as a result, Hunt's analysis is largely limited to positive cases—examining the pattern of editorials prior to wars or crises. In addition, his analysis is also limited by the reliance on domestic newspapers (as opposed to English-speaking in our case), so that a systematic analysis is particularly difficult. Finally, it is limited to interstate wars.

Finally, we note that this article is not about the impact of the media on foreign policy. We mean to look at the media as an indicator, not as a cause (see Strobel (1997) for a different perspective). This is not to mean that the media might not influence foreign policy, (Rosenau 1961); (Holsti 1996), but simply that our interest is in predicting war, not in making any causality claim about "who influences whom".

3 The Data

3.1 Measuring Tensions

To estimate domestic and international tensions, we relied on Google's database of newspapers, *Google News Archive*.[4] This wide collection includes a large proportion of all English-speaking newspapers, ranging from major publications such as *The New York Times*, *The Washington Post* or *The Guardian*, to more obscure local ones such as *California Oil Worker* or *The Cambridge City Tribune*.[5] In all, the database spans more than 200 years and consists of over 60 million pages. It also includes as subsets major providers of news archives such as Proquest Historical Newspapers, thereby making it the world's largest database in terms of the number of articles referenced. This comprehensiveness has the added advantage of smoothing out any particular newspaper's biases, such as those caused by their geographic location (Thai newspapers, say, might not have written as much about WWI as Germany's), their political orientation (conservative or liberal) or their substantive focus

[4] `http://news.google.com/newspapers?nid=t_XbbNNkFXoC`

[5] see http://news.google.com/newspapers for a partial list

(politics, economics or art).

Within this data, we searched the entire text of every article for every week from 1902 to 2011 (data prior to 1902 in Google's database was less reliable). We then counted the number of articles mentioning a given country, together with a set of keywords typically associated with tensions. The list of keywords, generated using a thesaurus to avoid any personal or linguistic bias, is the following: *tension(s), crisis, conflict, antagonism, clash, contention, discord, dissent, disunion, disunity, feud, division, fight, hostility, rupture, strife, attack, combat, shell, struggle, fighting, confrontation, impasse.* Thus, a sample search would be "France AND tensions OR crisis OR conflict [...]" for newspapers published between July 22^{nd} and July 29^{th}, 1914. This search yielded six results, indicating that six newspaper articles mentioned at least one of our keywords in their text. We repeated this procedure for every week from January 1^{st} 1902 to January 1^{st} 2011, and for every country included in the Correlates of War dataset (Correlates of War Project 2008).[6] The resulting dataset consists of 109 years worth of weekly time series for 167 countries, for a total of more than half a million data points to analyze.[7]

[FIGURE 1 ABOUT HERE]

Undoubtedly, the list remains ad hoc, and another set of keywords may better measure tensions or predict conflict. To ensure the robustness of our

[6] The list of countries includes only those that still exist as of today, and therefore excludes ones such as Austria-Hungary

[7] Note that some countries, such as those emerging from decolonization, appear only later in the sample.

results, we have therefore collected news counts for other sets of words.[8] Overall, we found no significant qualititative differences in the results. For example, using only *tensions* as a keyword led to a lower total number of news (since it excludes all news mentioning for example *conflict* but not *tensions*), but to a time series highly correlated with ours.

Nevertheless, our data remains imperfect. First, newspapers from English-speaking countries are over-represented in the archive and, in fact, we limit our analysis to articles written in English. This implies a bias toward English-speaking countries, which we account for by including an *English* dummy for the US, the UK and Australia in our models.[9]

Moreover, the total number of news articles is skewed toward more recent years—in particular since the 1980s. This is due to two main factors. First, the effort to collect and assemble a comprehensive newspaper database is long and costly, and has focused more on recent years than on the distant past. Second, lower production costs and rising population and education levels have increased readership. As a result of this increasing trend, it is more difficult to determine whether an increase in conflict-related news is due to the sheer increase in the total number of articles published, or to an actual increase in conflict-related concerns. We address this issue in our model by adding an interaction term between news count and years posterior to 1980.

[8]We did this only for a limited set of countries, given the computational challenges involved

[9]However, it does also include a minority of newspapers in English based in non-English speaking countries, such as the Japanese Daily Yomiuri.

Another difficulty is the potential influence that journalists might exert on each other. In fact, if they draw part of their information or article ideas from each other, cascades can emerge and lead to panics out of objectively minor events. However, the wide diversity of newspapers included in the data should attenuate this effect.

Moreover, our crude search does not allow us to determine who will fight with whom. We might for example know that both France and Germany are experiencing tensions, but not whether these tensions are in relation to one another, to a third country, or simply happen to spike at the same time for altogether different reasons. This problem could be addressed by searching for "War AND France AND Germany" instead of simply "war AND France". This approach would give us a more fine-grained view of tensions, but it also answers a different question: not only whether tensions are rising, but also with whom, and is therefore not applicable to civil wars. It also requires far more measurements, since 167 countries imply 16110 dyads, and hence more than 92 million weekly data points to collect—a technical challenge that we reserve for a future paper.

Another limitation of the data is that, although the entire text can be searched for specific keywords or sentences, legal access limitations imply that the content cannot be processed for more complex analysis. This implies a certain crudeness in our time series, in that they are limited to a simple count of articles mentioning specific terms, and cannot interpret the meaning of the article. Thus, "war will not occur" increases our estimate of tensions to the same extent as "war will occur". This limitation could be circumvented by relying on full-text databases such as Proquest Historical News, but their scope is far more limited than Google News Archive. Moreover, a newspaper contributor writing about her belief that a conflict will *not* break out still reveals existing concerns that need to be dispelled and, as such, *should* be treated as a sign of tensions.

The crudeness of our data collection also implies that we fail to pick up isolated but prescient accounts. For example, a single journalist newspaper predicting war earlier than the others will be largely inconsequential, even if its analysis is particularly compelling. By looking at sheer numbers, we do not take into account the quality of writer's analysis. However, the same could probably be said of any analysis performed without the benefit of hindsight, since identifying prescient accounts is almost impossible.

Another difficulty is caused by propanganda, which can bias news reports, but the fact that we use worldwide news articles published in English avoids a large part of this problem, since American, British and Australian newspapers are less likely to be subjected to the editorial pressures of, say, the French or German governments.[10]

Finally, we noted some marginal mistakes in Google's article dating algorithm. For example, some articles referencing July 28 1914 were listed as having been published in 1914, even though they were published much later.

[10]Of course, this problem is not solved for the wars that do involve English-speaking countries but even then, the wide range of newspapers and of countries of origin should to a large extent smoothen out any potential bias or propaganda efforts.

This implies a slight bias in the data, since the news count on the day of the outbreak will be overestimated. However, these errors are rare and cannot in any significant way account for our results. Moreover, our interest is in the pattern leading to conflict—not in the day of its outbreak.

While the data could certainly could be improved in the future, it is to our knowledge the most comprehensive, systematic and uniform estimate of tensions throughout countries, and we show that the simple count we rely on can already produce substantial results.

3.2 Conflicts

Conflicts can be broadly categorized according to their scale and the actors involved. First, conflicts range from aggressive speeches to the simple display of force to full-scale wars with thousands of deaths. Second, they can involve either only states ('interstate conflicts'); one state against rebel groups (civil war or 'intrastate'); or states with non-state armed groups with no defined territorial base (extra-state wars) (Sarkees & Wayman 2010*b*).

We study the rise of tensions for all militarized conflicts included in the Correlates of War (CoW) or the MID data (Sarkees & Wayman 2010*b*, Faten, Glenn & Stuart 2004).[11] This includes all inter-, intra- and extra-state conflicts recorded with a starting date of January 1902 to December 2010 (see table 1 for a breakdown). Wars prior to 1902 are excluded because Google's

[11] Intrastate war data v4.1, interstate war data v4.0 and version 3.0 of the Extra-State War data set; MID v3.0.

news collection is not as reliable and complete in the 19^{th} century. Note also that we use country-conflicts, which implies that World War I includes an entry for all 15 participants listed in the CoW data. In total, this means that our dependent variable is composed of 4,396 militarized interstate disputes and 223 wars with at least 1,000 military casualties (95 interstate, 98 intrastate and 30 extra-state, for a total of more than 500,000 entries

[TABLE 1 ABOUT HERE]

This variety will test the ability of our measure of tension to announce not only large-scale interstate wars, but also bloodless domestic conflicts. Moreover, we are interested in the *types* of war that are best anticipated, which justifies our inclusion of all types of war—are civil wars, for example, more or less predictable than interstate wars? We show in ongoing work that our results extend to dyadic data (replicating work by Beck, King & Zeng (2004) with the addition of the conflict-related news variable), but we use monadic data here to understand the type of war that is better understood.

4 Conflict-related News Signal Geopolitical Risk.

4.1 Bivariate Relationship

The average weekly number of conflict-related news (mode = 0, mean = 28.28, median = 6, sd = 126.92) varies considerably in time—for example increasing in the United States from an average of 28 in the 1900s to 512 in the 2000s—and space—ranging from 0.23 for Suriname to 249 for Iraq. However, conflict-related news counts dramatically increase in the months and years preceding conflict, and rapidly recede thereafter (while remaining higher than average long after the war) (Fig. 1).

[FIGURE 1 ABOUT HERE]

This trend applies independently of war size (casualties) or type (inter- or intra-state). The pattern emerges remarkably early: a visible upward trend appears at least three to five years before large wars, and two to four years before MIDs. Note that large wars ($> 10,000$ deaths) have a slower downward curve following war, which is not surprising since they last longer and are costlier, and hence their impact drags on longer than MIDs, which involve fewer casualties and are limited in duration. We also find, not surprisingly, that the number of conflict-related news is much higher within the year that precedes the outbreak of war (9) than at other times (4), for wars of any scale or type (table 2).

[TABLE 2 ABOUT HERE]

Yet, while the bivariate relationship between news and time to war suggests that conflict-related news counts vary significantly with the proximity to war, it is hardly sufficient evidence of a causal relationship. Indeed, it may be that the number of conflict-related news simply reflects changes in other variables (e.g., military spending), and hence does not carry additional information. Moreover, it does not inform us about the evolution of the number of conflict-related news in cases where war does *not* occur. We therefore tested the specific explanatory power of news with a logistic regression model in which, in addition to conflict-related news counts, we included potentially confounding variables that have been identified as significant in the literature on conflict.

4.2 Multivariate Analysis

We fitted the following standard logit model:

$$\mathbb{E}[Y_{it}|\mathbf{X_{it}}] = 1/(1 + e^{-\boldsymbol{\beta} X_{it}}),$$

where $Y_{it} \in \{0, 1\}$ is the occurrence or not in country i and at time t of a war within the next year, $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_k]$ is a vector of coefficients, and $\mathbf{X_{it}}$ is a matrix of explanatory variables which, in addition to conflict-related news, includes other variables that are likely to affect a country's odds of experiencing a conflict: a variable measuring a country's national material capabilities using the Composite Index of National Capability (*Cinc*) from the Correlates of War (Singer, Bremer & Stuckey 1972).[12] We also included

[12]NMC v. 4.0. We time-lagged Cinc by one year, because its yearly estimation would seriously bias our result. Consider for example a war starting on January 31^{st}. Our prediction on January 1^{st} of that year would then include the Cinc data for that year (which will be very high, given the probable increase in military spending after the start of the war), which is information that was not available in January of that year.

a variable measuring the yearly change in a country's Cinc to account for the explanation that large and rapid shifts in power can lead to war (Organski & Kugler 1981, Chadefaux 2011). The effect of political regime types (e.g., democracy vs. autocracy) on the probability of war has also been the subject of much research (Ray 1998). We therefore also included a *Polity* variable (Marshall, Jaggers & Gurr 2002).[13]

Finally, we included various lagged and interaction variables. First, we expect a rising trend in reports about tensions to be more indicative of a coming conflict than a decreasing one. We therefore added variables measuring the evolution of news counts from one month to three years prior to war. Second, the number of news written about tensions or war is itself probably not independent of other variables, which is why we added interaction terms. For example, whether a country is English-speaking (e.g., the US), its power (Cinc) or its regime type (Polity) all probably increase, all else constant, the number of conflict-related news that a country receives. Similarly, an ongoing war (*War ongoing*), a recent MID (*Days since MID*), or a large number of past conflicts (*N past wars* and *N past MIDs*) would most likely boost the count. We also included the total number of news in the world (*World news*) to control for the fact that increases in conflict-related news in one country do not necessarily indicate an increase in tensions, but might simply reflect a general trend. We also added a dummy for years after 1980 (> 1980) to account for the dramatic increase in overall news in the following decades.

[13] For the same reasons as for Cinc, we actually use the Polity value lagged by one year.

Finally, we included a variable measuring the number of years since the last conflict (Peace Years) as a measure of temporal dependence (Beck, Katz, and Tucker 1998). This variable controls for the possibility that conflict is more likely to erupt after previous disputes than after a long period of peace.[14]

[FIGURE 2 ABOUT HERE]

We tested several models. Model 1 is the baseline model, including only a constant and the 'time since last conflict' variable. Model 2 is adds the weekly number of conflict-related news to model 1, but no control variables. Model 3 is the 'structural' model, including variables that have been found to be important predictors of conflict in the literature (but not conflict-related news). Model 4 includes all control and interaction variables described above.

The results of the multivariate model confirm the results of the simple correlation we described above. We found that the risk of war increases significantly with the number of conflict-related news (figure 2), even after controlling for the structural variables described above. This result applies whether we define conflict as those with more than 10,000 battle deaths; those with deaths between 1,000 and 10,000; or those with less than 1,000 deaths. It also holds regardless of war type, for interstate, intrastate and extrastate wars alike (tables 3 and 4).

[14] As Beck, Katz, and Tucker show, the addition of this variable turns an ordinary logit analysis into a grouped proportional hazard model (Beck, King & Zeng 2004).

5 Predictive Power of Conflict-Related News

5.1 Deriving a weekly risk-index

While an increased number of conflict-related news signals a higher risk of war, forecasting wars remains a needle in a haystack problem. Large-scale wars, for example, occur in only about every 7,700 observations (table 1). While the addition of the conflict-related news variable significantly improves the fit of the model, a better test of the value added of our measure of tensions is its ability to improve predictions. That is, can we estimate the probability of a coming war (out of sample) using the measure of tensions we derive from newspapers better than without?

Our predictions are based on a risk-index, which we derived recursively for every country and every week, using only information information available at the time. For example, we used all information available up to January 1^{st} 1914 to estimate the coefficients of our logistic model, and applied these coefficients to the value of the independent variables on January 1^{st}, thereby obtaining a risk-index for that week. We repeated this procedure for every week from 1920 to 2010 (previous years were used for learning).[15]. This way, we recursively estimated our model and derived a risk-index—an estimate of the probability to experience conflict—for every country-week.

The predictive power of the resulting risk-index can be evaluated along

[15] More precisely, we use information about conflict up to January 1^{st}, 1914, and independent variable data up to Jan 1^{st} 1913, since our dependent variable is the occurrence of a conflict within the next twelve months

two main dimensions: first, its calibration—the ability to assign subjective probabilities to outcomes that correspond to their objective probability. Thus, events with an estimated predicted probability of 20% should occur about 20% of the time; second, its discriminating power—the ability to assign a higher probability to outcomes that occur than to those that do not. A model with good discrimination will assign higher risk values to countries that are going to experience a war than those that will remain peaceful. It is necessary to use both measures, as a model may have strong calibration but weak discrimination, or vice-versa.[16]

5.2 Improved Binary Predictions

An important test for the predictive power of the risk-index derived is its ability to correctly answer the question "Will war happen next week?". We therefore made for every week a prediction about peace or war for the coming week, in each country, based on the value of the risk-index derived for that week. A warning was issued ($\hat{Y} = 1$) when that week's risk-index crossed a given threshold. An alarm raised when war did not actually occur in the following week (i.e., more than 7 days, but no more than 15 days ahead) was counted as a false positive (FP). On the contrary, failure to raise an alarm when war did happen in the next week was counted as a false negative

[16] For example, a model which estimates the risk to be 49% prior to all peace events, and 51% prior to all war events, has perfect discrimination, but poor calibration. On the contrary, a model that assigns to all events a probability equal to the prevalence of the outcome (2% here) has perfect calibration, but no discrimination.

(FN). An alarm in the week prior to the warning time is a true positive (TP). Clearly, the choice of the threshold implies a trade-off between the prediction of a large proportion of all wars (low threshold) and the avoidance of false alarms (high threshold)(Fig. 3a).

For major wars, for example, we found that we can choose a threshold such that a warning is correctly issued one week ahead of 71.23% of wars ($P(\hat{Y} = 1|Y = 1) = .7123$), while false warnings occur in 16.77% of cases ($P(\hat{Y} = 0|Y = 1) = .1677$). If false alarms are deemed too costly, policy makers may instead prefer a stricter threshold, for example one such that a warning is issued ahead of 42.47% of wars, and false warnings represent only 5.4% of predictions. This compares very favorably to a model without our measures of tensions which, to maintain the same level of false positives, could only predict 30.77% of conflict or, to maintain the same level of true positives, would almost double our false positive rate (9.0%).

[FIGURE 3 ABOUT HERE]

The area under the curve (AUC) summarizes the predictive power of our index for all values of the threshold (Fig. 3a). It can be interpreted as the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative one (Bamber 1975). Thus, an AUC of 0.808 for large-scale wars means that in 80.8% of weeks immediately preceding a large war, the value of our index was larger than at other times. This represents an improvement of about 56% in predictive power over a model that does not include measures of tensions (AUC = 0.777), as compared to a baseline model that includes only a 'time since the last conflict' variable (AUC = 0.722) (Fig. 3b) (Beck, King & Zeng 2004). The improvement is even more pronounced for interstate wars (191% improvement) and for all conflicts in general (100% improvement). Overall, the inclusion of conflict-related news significantly improves the predictions over a model with only structural variables for all types of wars.[17] In absolute terms, however, conflicts with lower casualties, are harder to predict (AUC = 0.727 for small wars and 0.742 for all conflicts), and interstate conflicts are better predicted than civil wars (AUC = 0.767 and 0.668 respectively).

Finally, forecasting wars is most useful if it can give policy makers sufficient warning time to react, and ideally avert the disaster. We therefore tested the ability of various models to correctly answer the question "will war occur in exactly x months?". That is, we increased the warning time from one week to x months, and asked each week whether a war would start x months ahead. Here too, we found that our risk-index significantly outperforms other models, even with a warning time of more than one year. Indeed, we see in figure 4 that the AUC for the model with conflict-related news is higher than for other models not just for a short warning time (one week), but also well before the onset of war.

[17]The reader might wonder about the discrepancy between our AUC for interstate and those obtained by, for example, Beck, King & Zeng (2004). The different results are explained first by different dependent variables (dyadic wars vs. country-wars), as well as by differences in the coarseness of our respective time series (yearly for Beck, King & Zeng (2004), weekly data here).

[FIGURE 4 ABOUT HERE]

Of course, the improvement diminishes with the time to the conflict: 3 years ahead, newspapers offer little insight beyond what we could assess from structural factors, but the closer we get to the conflict onset, the higher the improvement provided by an analysis of news. This suggests that newspapers provide valuable information beyond structural factors (e.g., military spending) more than one year ahead of conflict (Fig. 4). The warning time is shorter for intrastate wars (about 6 months), than for interstate wars (more than one year). Interestingly, we find that while major wars are more reliably predicted than smaller conflicts in absolute terms (higher AUC), the added value of conflict-related news begins much later (closer to conflict) than for smaller conflicts. We conjecture that large wars may be more predictable using structural variables, and hence that news provide less additional information well ahead of conflict. Only as conflict becomes closer do news provide improved predictions, as they are better able to "time" the onset. On the contrary, small skirmishes may not reflect any structural risk, and hence can be overseen by a structural analysis but might still be the subject of early media attention.

5.3 Improved Probabilistic Predictions

Policy-makers are not only interested in binary predictions—will war occur?—but also in estimates of the probability of an event: "What is the probability of a war onset next week?". We show that, in addition to improving binary predictions, the tension-based index also yields improved estimates of the probability of war. Just as we derived predictions recursively, we also calibrated our model every week, using information about the past performance of our index. We calibrated the risk-index recursively by classifying our predictions into 10 categories (risk-index of [0-10)%, [10-20)%, etc.), and comparing these predictions to the average outcome following these predictions. In particular, we assessed each week the extent to which our past predictions had been over or under-optimistic, and adjusted our current estimation accordingly. If, for example, past risk-indices had estimated a 50% chance of conflict within the coming year, but it turned out that conflict occurred in only 30% of these cases, we concluded that we had overestimated the probability of war, and therefore adjusted our future estimates from 50% to 30% (but not estimates in other categories). This adjustment was performed recursively every week using information about the past performance of the index.

We found that the occurrence of a conflict within a one-year window can be forecasted with up to 85% confidence, meaning that wars occurred within one year of about 85% of cases in which we predicted an 85% risk. Figure 5 shows the agreement between the estimated risk derived from the model and the actual risk of war for different time windows (occurrence of war within the next x months). Better calibrations are those that (i) follow the 45^0 line and (ii) predict a large range of values. A smaller time window makes it difficult to predict events with high certainty, given the lower probability of an event occurring within that time frame. Conversely, peace tended to prevail when our index forecasted a low risk of conflict.

[FIGURE 5 ABOUT HERE]

While it is difficult to evaluate how good these forecasts are in absolute terms, we can at least show that they represent significant improvements over models that do not include information about tensions (model 3). Various indicators have been proposed in the literature for that purpose. The Brier score measures the mean squared error of the probability forecast. It is defined as $BS = \frac{1}{N}\sum_{t=1}^{N}(p_i - Y_i)^2 \in [0,1]$, where p_i denotes a prediction made and Y_i be the occurrence or not of a conflict within the following 12 months. A lower score represents higher accuracy (Brier 1950). The results in table 5 show that our predictions are on average significantly closer to the actual occurrence than a model that does not include a measure of tensions. The F1 score is based on the precision p (the percent of correct positive predictions) and recall r (the percentage of positive cases detected) measures: F1 $= \frac{2pr}{p+r} \in [0,1]$. A higher value indicates better predictive power. Finally, Shapiro's Q is an overall measure of the accuracy of the model, incorporating both calibration and discrimination. Q is calculated as $\sum_{i=1}^{N}\log(2p_i) + \sum_{i=1}^{N}(1-Y_i)\log(2(1-p_i))$. A higher Q value indicates a better prediction.

[TABLE 5 ABOUT HERE]

We see in table 5 that all measures favor the model with conflict-related news.

6 Conclusion

The prediction of wars has received relatively little interest in the literature. This is in sharp contrast with finance, a field in which prediction is very difficult, yet has occupied many researchers. One important difference between the two fields used to be the availability of data: financial data is readily available in fine-grained time-series, whereas information about military spending, diplomatic agreements and other international events is far more difficult to collect, harmonize and analyze.

In this context, the present paper intended to make three main contributions. First, we collected a new dataset on the weekly occurrence by country of certain conflict-related terms. While imperfect—the list is ad hoc and is a simple count, not an in-depth analysis of the content of each article—this data offers some rare advantages. Its frequency (weekly) is far superior to most existing data. Its time-span is also long, going back to the beginning of the 20^{th} century. Finally, it is largely independent of harmonization issues, reliability or manipulation. The estimated tensions derived can therefore be a valuable tool for future research. In particular, numerous questions in international politics remain unsettled for lack of appropriate data. For example, are democracies better at resolving conflict short of war, or at avoiding tensions with one another altogether? How do tensions spread, geographically and through networks (e.g., alliances)? Instead of relying on rare binary events, fine-grained measures of tensions allow us to study the immediate impact of certain changes or shocks, without needing to draw difficult connexions between distant events.

Our second contribution is to show that this data is a strong predictor of conflict. The number of conflict-related news increases dramatically prior to conflicts, and therefore we can conjecture that contemporaries do witness and notice the rise of tensions. Wars rarely rarely emerge out of nowhere, though more research will be needed on the interesting cases in which journalists failed to pick up relevant clues, and hence where war came as a surprise.

Finally, we showed the ability of our measure of tensions to function as a reliable early warning signal, using only information available at the time. In particular, it improves for every type of war (inter- or intrastate, large or small) the precision with which we can answer questions such as "Will a war occur next year?", "What is the probability of a war happening next year?" and "Will a war happen in exactly one year from today?".

Overall, these findings demonstrate the importance of developing additional fine-grained measures of geopolitical tensions, as they help develop well-calibrated and reliable risk indices, and provide policy-makers accurate and early warnings for war. For all the limitations of the data, the improved predictions generated from its addition is encouraging. We hope that this work will trigger attempts to collect better data along these lines, including perhaps the use of content analysis and the extension to news in other languages.

References

Azar, E.E. 1980. "The conflict and peace data bank (COPDAB) project." *Journal of Conflict Resolution* 24(1):143–152.

Bamber, D. 1975. "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph." *Journal of mathematical psychology* 12(4):387–415.

Beck, N., G. King & L. Zeng. 2004. "Theory and evidence in international conflict: a response to de Marchi, Gelpi, and Grynaviski." *American Political Science Review* 98(2):379–389.

Beck, Nathaniel, Gary King & Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–36.

Brier, G.W. 1950. "Verification of forecasts expressed in terms of probability." *Monthly weather review* 78(1):1–3.

Chadefaux, T. 2011. "Bargaining over power: when do shifts in power lead to war?" *International Theory* 3(02):228–253.

Choucri, N. 1974. "Forecasting in international relations: problems and prospects." *International Interactions* 1(2):63–86.

Correlates of War Project. 2008. "State System Membership List, v2008.1." Downloaded from http://correlatesofwar.org.

De Marchi, S., C. Gelpi & J.D. Grynaviski. 2004. "Untangling neural nets." *American Political Science Review* 98(2):371–378.

Deutsch, K.W. 1957. "Mass communications and the loss of freedom in national decision-making: a possible research approach to interstate conflicts." *Conflict Resolution* 1(2):200–211.

Faten, Ghosn, Palmer Glenn & Bremer Stuart. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21(2):133–154.

Fearon, J.D. & D.D. Laitin. 2003. "Ethnicity, insurgency, and civil war." *American Political Science Review* 97(1):75–90.

Ferguson, Niall. 1999. *The Pity of War*. Basic Books.

George, A.L. 1956. "Prediction of political action by means of propaganda analysis." *The Public Opinion Quarterly* 20(1):334–345.

George, A.L. 1959. *Propaganda Analysis: A Study of Inferences Made from Nazi Propaganda in World War II*. Row, Peterson and Company.

Glaser, C.L. 2000. "The causes and consequences of arms races." *Annual Review of Political Science* 3(1):251–276.

Gleditsch, K.S. & M.D. Ward. 2011. "Forecasting is Difficult, Especially the Future: Using Contentious Issues to Forecast Interstate Disputes.".

Gochman, C.S. & Z. Maoz. 1984. "Militarized interstate disputes, 1816-1976." *Journal of Conflict Resolution* 28(4):585–616.

Holsti, O.R. 1963. "The value of international tension measurement." *Journal of Conflict Resolution* 7(3):608–617.

Holsti, O.R. 1996. *Public opinion and American foreign policy*. Cambridge Univ Press.

Hunt, W. B. 1997. *Getting to war: Predicting international conflict with mass media indicators*. University of Michigan Press.

Huth, P.K. 1998. *Standing your ground: Territorial disputes and international conflict*. University of Michigan Press.

Köhler, G. 1975. "Le problème de la définition d'une échelle d'hostilité pour la mesure des tensions." *Études internationales* 6(1):30–46.

Lasswell, H.D. 1971. *Propaganda Technique in World War I*. MIT press.

Lebovic, J.H. 1998. "Consider the source: Organizational bias in estimates of foreign military spending." *International Studies Quarterly* 42(1):161–174.

Lebovic, J.H. 1999. "Using military spending data: The complexity of simple inference." *Journal of Peace Research* 36(6):681–697.

Leetaru, K. 2011. "Thread: Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space." *First Monday* 16(9-5).

Leng, R.J. 1987. *Behavioral Correlates of War, 1816-1979*. Inter-university Consortium for Political and Social Research.

Marshall, Monty G., Keith Jaggers & Ted Robert Gurr. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010. Version p4v2010." *Center for International Development and Conflict Management* .
URL: *http://www.cidcm.umd.edu/inscr/polity/index.htm*

McClelland, C.A. 1984. *World Event/Interaction Survey (WEIS) Project, 1966-1978*. Inter-university Consortium for Political and Social Research.

Newcombe, A.G., N.S. Newcombe & G.D. Landrus. 1974. "The development of an inter-nation tensiometer." *International Interactions* 1(1):3–18.

Organski, A.F.K. & J. Kugler. 1981. *The war ledger*. University of Chicago Press.

Powell, Robert. 2004. "The inefficient use of power: Costly conflict with complete information." *American Political Science Review* 98(02):231–241.

Ray, J.L. 1998. "Does democracy cause peace?" *Annual Review of Political Science* 1(1):27–46.

Rosenau, J.N. 1961. *Public opinion and foreign policy: An operational formulation*. Vol. 35 Random House New York.

Sarkees, Meredith Reid & Frank Wayman. 2010*a*. *Resort to arms: International and civil wars, 1816-2007*. CQ Press.

Sarkees, M.R. & F. Wayman. 2010*b*. *Resort to War, 1816-2007*. CQ Press.

Singer, David J., Stuart Bremer & John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills: Sage pp. 19–48.

Strobel, W.P. 1997. *Late-breaking foreign policy: The news media's influence on peace operations*. United States Inst of Peace Press.

Surowiecki, J., M.P. Silverman et al. 2007. "The wisdom of crowds." *American Journal of Physics* 75:190.

Taleb, N.N. 2011. *The black swan: The impact of the highly improbable*. Random House Inc.

Tetlock, P. 2005. *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Weidmann, N.B. & M.D. Ward. 2010. "Predicting conflict in space and time." *Journal of Conflict Resolution* 54(6):883–901.

A Figures

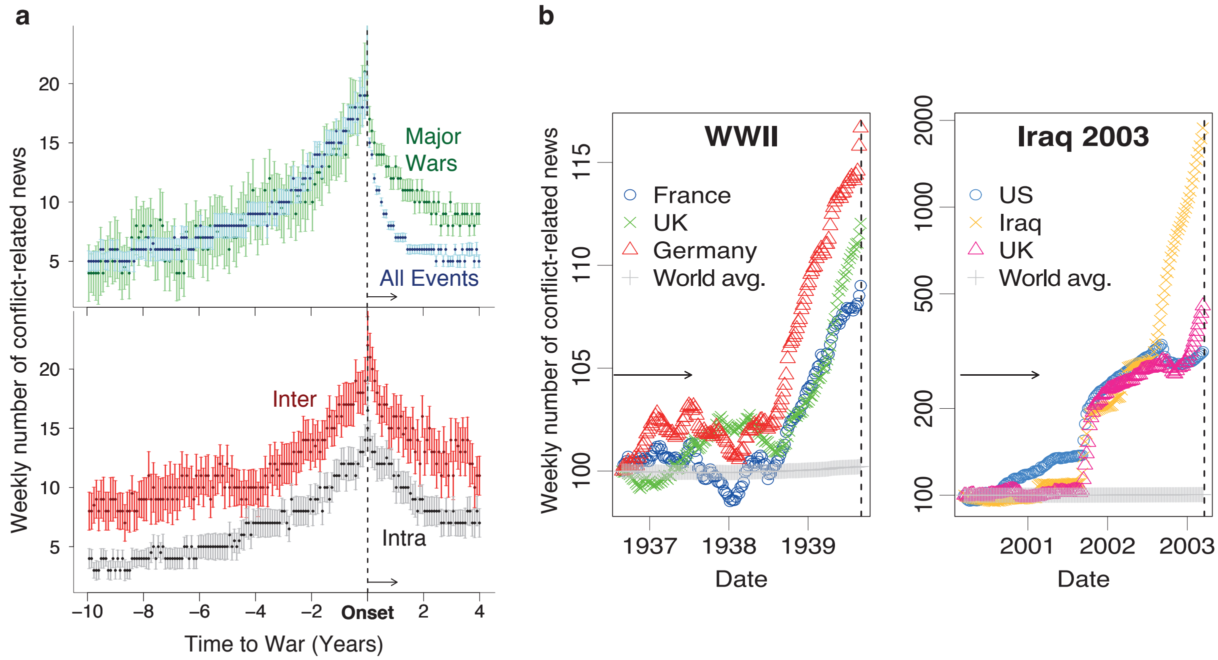

Figure 1: Median weekly number of conflict-related news as a function of time to conflict. **a.** Evolution prior to all 68 conflicts of any type (inter-, intra- or extra-state) with at least 10,000 battle deaths (green), and prior to all 4,530 conflicts of any size or type (blue); **b,** evolution prior to all interstate wars (red) and intrastate wars (black) with at least 1,000 battle deaths. The vertical bars represent the standard error of the median

. **b.** Number of conflict-related news prior to WWII and the second Iraq war, for selected countries.

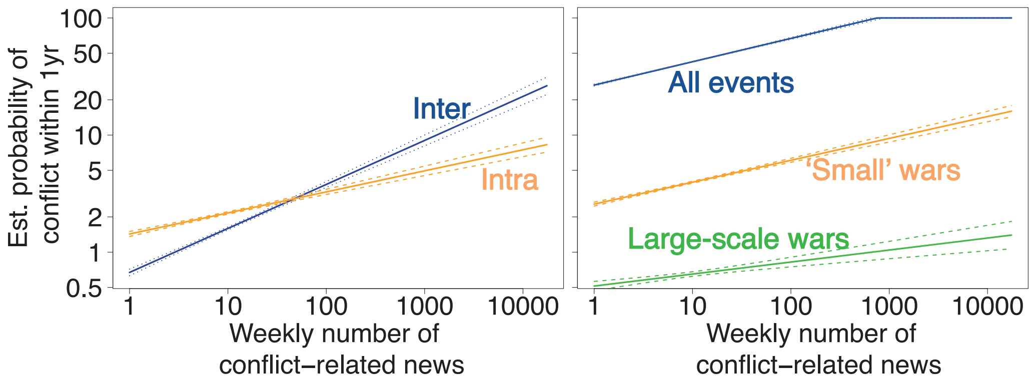

Figure 2: Estimated probability (%) of conflict onset within one year, as a function of the present (weekly) number of conflict-related news.

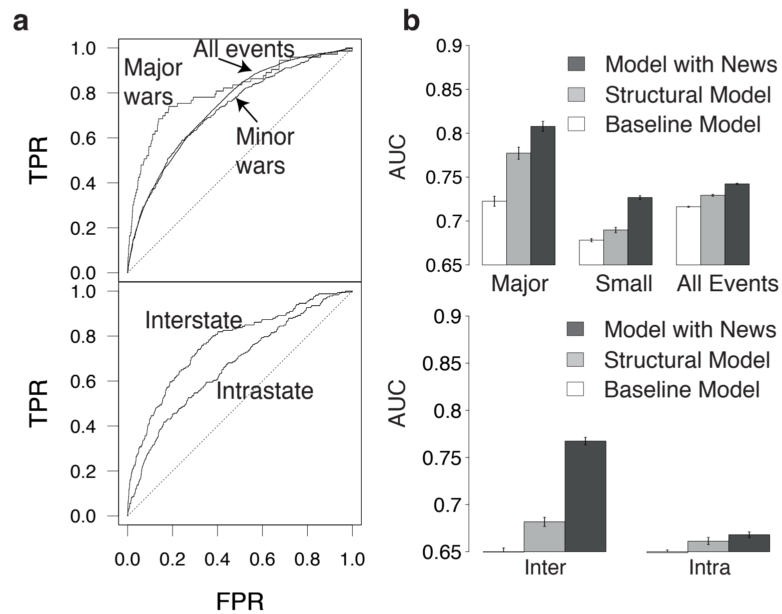

Figure 3: Relative predictive performance of the tension-based index. **a,** Receiver operating characteristic (ROC) curves displaying the trade-off between the true positive rate (TPR $\equiv P(\hat{Y} = 1|Y = 1)$) and the false positive rate (FPR $\equiv P(\hat{Y} = 1|Y = 0)$) for a warning time of one week. The curves show the trade-offs in the choice of a threshold for the detection of war, between the detection of a large proportion of wars (high TPR) and false warnings (high FPR). **b,** Area under the curve (AUC) of three models for different classes of war, for a warning time of one week. The error bars represent the standard errors of the AUC, obtained by bootstrapping.

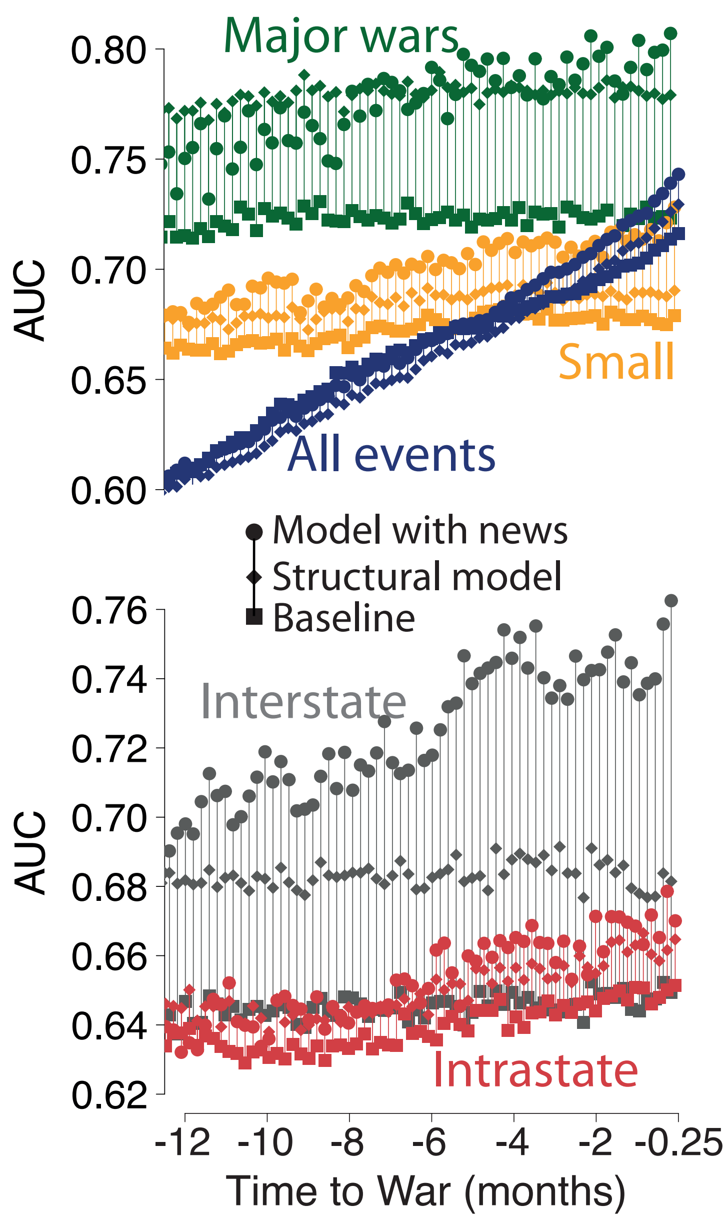

Figure 4: Predictive power of the three models as warning time increases, as measured by the area under the curve.

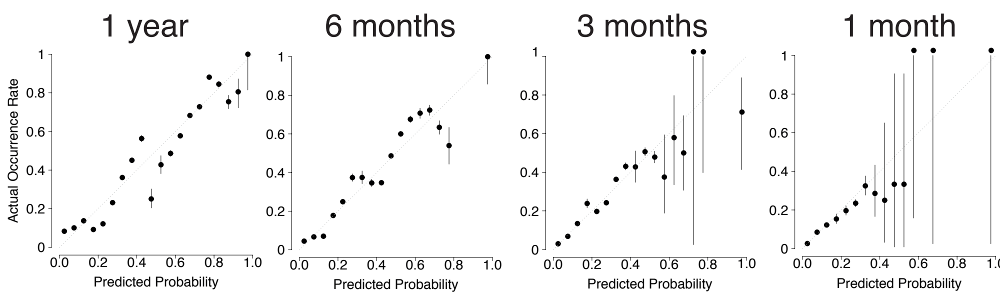

Figure 5: Probability of conflict within one year: calibration of the risk index against actual occurence rate of conflict, 1920–2011. Vertical bars represent confidence intervals.

B Tables

Type	**Battle deaths**			
	<1,000 (‘minor’)	**[1-10,000) (‘small’)**	**10,000+ (‘major’)**	**Total**
Inter	4,193	46	22	4,261
Intra	119	69	39	227
Extra	8	27	7	42
Total	4,320	142	68	4,530

Table 1: Frequencies of Country-Events (1902–2010) by conflict type and battle death count. A country-event is coded as one for every week in which a conflict breaks out, 0 otherwise

Conflict Type	W	$\overline{W}$	**U**	n_1	n_2	**P (one-tailed)**
All Events	9	4	2.331×10^{10}	$112,831$	$352,641$	< 0.001
Small Wars	11	5	5.168×10^{9}	$19,333$	$449,229$	< 0.001
Large-scale	15	6	1.376×10^{9}	$4,476$	$498,333$	< 0.001
Interstate wars	16	6	3.290×10^{9}	$10,287$	$498,912$	< 0.001
Intrastate wars	11	6	3.174×10^{9}	$11,332$	$476,971$	< 0.001

Table 2: Mann-Whitney U statistic comparing the median number of conflict-related news within the year preceding war (W) to its median during weeks not within one year of a conflict ($\overline{W}$), for various categories of conflict. We use the Mann-Whitney U statistics instead of the typical t-test because of the skewed distribution of the number of conflict-related news. However, similar results hold with the t-test.

	Model 1		Model 2		Model 3		Model 4	
Const.	-0.38^{***}	(0.00)	-0.74^{***}	(0.01)	-0.84^{***}	(0.01)	-1.76^{***}	(0.02)
Peace Years	-0.00^{***}	(0.00)	-0.00^{***}	(0.00)	-0.00^{***}	(0.00)	-0.00^{***}	(0.00)
News (log)			0.15^{***}	(0.00)			0.69^{***}	(0.01)
N past MIDs					0.01^{***}	(0.00)	0.01^{***}	(0.00)
N past wars					0.00^{***}	(0.00)	-0.01^{***}	(0.00)
War ongoing					0.42^{***}	(0.01)	0.38^{***}	(0.02)
English					-0.51^{***}	(0.03)	-2.27^{***}	(0.15)
Cinc (lag)					0.06^{***}	(0.00)	0.11^{***}	(0.01)
Power Shift					-0.16^{***}	(0.01)	-0.12^{***}	(0.01)
Polity (lag)					-0.02^{***}	(0.00)	-0.02^{***}	(0.00)
> 1980					-0.10^{***}	(0.01)	0.11^{***}	(0.02)
$\Delta \mathrm{News}_{t-1mo}$							-0.04^{***}	(0.01)
$\Delta \mathrm{News}_{t-3mo}$							-0.04^{***}	(0.01)
$\Delta \mathrm{News}_{t-6mo}$							-0.02^{**}	(0.01)
$\Delta \mathrm{News}_{t-1yr}$							-0.04^{***}	(0.01)
$\Delta \mathrm{News}_{t-2yr}$							0.03^{***}	(0.01)
$\Delta \mathrm{News}_{t-3yr}$							0.00	(0.01)
World news							0.28^{***}	(0.01)
$> 1980 \times$ news							-0.00	(0.01)
logNews $\times$ mvaCincLag365							-0.02^{***}	(0.00)
logNews $\times$ warOngoing							-0.03^{***}	(0.01)
World news $\times$ news							-0.26^{***}	(0.00)
English $\times$ news							0.42^{***}	(0.04)
Nagelkerke R-sq.	0.13		0.14		0.18		0.22	
Likelihood-ratio	43870.43		48017.21		64442.30		77477.77	
Log-likelihood	-259065.10		-256991.71		-248779.17		-242261.43	
Deviance	518130.20		513983.42		497558.33		484522.86	
AIC	518134.20		513989.42		497578.33		484568.86	
BIC	518156.36		514022.67		497689.14		484823.72	

Table 3: Logit model regressing the onset of conflict of any size or type within the next year against various variable combinations. The model with conflict-related news reported in the main text is model 4. The 'structural' model is model 3. The baseline model uses only a 'days since the last conflict' variable (model 1).

	β_{CRN}	**AIC baseline model**	**AIC model without CRN**	**AIC model with CRN**	**LR test**
Any size/type	0.69*** (0.01)	518,134	497,578	484,568	< 0.001
Deaths $\in$**[1k-10k)**	0.38*** (0.01)	179,511	172,621	169,501	< 0.001
Deaths $>$ **10k**	0.36*** (0.03)	51,292	47,022	45,056	< 0.001
Inter- ($>$ 1**k deaths)**	0.60*** (0.02)	94,019	89,172	86,269	< 0.001
Intra- ($>$ 1**k deaths)**	0.36*** (0.02)	111,648	108,794	107,515	< 0.001
Extra- ($>$ 1**k deaths)**	0.40*** (0.08)	25,040	18,528	17,993	< 0.001

Table 4: Results of logit models using different dependent variables. The first column reports the coefficient for the "News" variable. Thus, the top left number (0.69) corresponds to the bolded cell in table 3. The next three columns report the AIC statistic for the three main models (model 1, 3 and 4) in table 3. The last column reports the likelihood ratio tests of the model including conflict-related news (model 4 in table 3) against the same model without conflict-related news. The likelihood ratio test is a method for hypothesis testing by which the fit of a data set to a more complex model is compared with its fit to a simpler model using the likelihood ratio statistic (twice the ratio of the likelihoods of the two models). The improvement in fit is evaluated using a χ^2 distribution.

Test	Event Type	CRN	Struct.	Base
Brier	All Events	0.17278	0.17581	0.17864
	1k-10k	0.04284	0.04323	0.04434
	10k+	0.00837	0.00842	0.00887
	Inter	0.01860	0.01873	0.01936
	Intra	0.02440	0.02443	0.02451
F1	All Events	0.46013	0.45260	0.45517
	1k-10k	0.1365	0.12854	0.11681
	10k+	0.03628	0.03481	0.02867
	Inter	0.06417	0.05700	0.04835
	Intra	0.06892	0.06781	0.06256
Shapiro's Q	All Events	0.16658	0.16023	0.14531
	1k-10k	0.51486	0.51156	0.50400
	10k+	0.6477	0.6472	0.64111
	Inter	0.59974	0.59827	0.59577
	Intra	0.57653	0.57336	0.57313

Table 5: Measures of calibration of the risk-index. The model with the best predictive power is highlighted.