

# Conflict forecasting and its limits

Thomas Chadefaux

*Department of Political Science, Trinity College Dublin, 2–3 College Green, Dublin 2, Ireland*  
*E-mail: [thomas.chadefaux@tcd.ie](mailto:thomas.chadefaux@tcd.ie); ORCID: <https://orcid.org/0000-0002-8456-8124>*

**Editor:** Tobias Kuhn (<https://orcid.org/0000-0002-1267-0234>)

**Solicited reviews:** Thomas Maillart (<https://orcid.org/0000-0002-5747-9927>); Håvard Møkleiv Nygård (<https://orcid.org/0000-0002-8909-6146>); Nils Metternich (<https://orcid.org/0000-0001-8757-0409>)

Received 20 February 2017

Accepted 3 March 2017

**Abstract.** Research on international conflict has mostly focused on explaining events such as the onset or termination of wars, rather than on trying to predict them. Recently, however, forecasts of political phenomena have received growing attention. Predictions of violent events, in particular, have been increasingly accurate using various methods ranging from expert knowledge to quantitative methods and formal modeling. Yet, we know little about the limits of these approaches, even though information about these limits has critical implications for both future research and policy-making. In particular, are our predictive inaccuracies due to limitations of our models, data, or assumptions, in which case improvements should occur incrementally. Or are there aspects of conflicts that will always remain fundamentally unpredictable? After reviewing some of the current approaches to forecasting conflict, I suggest avenues of research that could disentangle the causes of our current predictive failures.

**Keywords:** Conflict, war, forecasting, tournaments, predictability

The literature on conflict and peace research has evolved over the past decades into a coherent field of inquiry (Van Holt et al. [58]). Of increasing interest have been forecasts of political events, and in particular violent ones such as international conflicts (Beck, King, and Zeng [2]), civil wars (M.D. Ward, Greenhill, and Bakke [60]), coups (Goldstone et al. [21]), or mass killings (Ulfelder [57]; Rost [40]). This is a welcome change in a field which, until recently, focused almost exclusively on explanation rather than prediction. This growing interest in forecasting in the academic community has been matched by increasing expectations from the military and the intelligence communities, who are working closely with academics to avoid some of the past intelligence failures and misestimations of the costs and risks of war.

Yet, despite our improving ability to forecast conflicts – thanks, in particular, to the growth of big data and new methods – our expectations should be balanced with a healthy dose of skepticism about what is and is not possible to predict. We know that our ability to forecast conflicts is limited because of imperfect models and data (Cederman and Weidmann [11]). But what we do not know is the extent to which wars are predictable at all. Are there inherent limitations to our ability to predict the onset and evolution of conflicts? And more generally, is the world of international relations and conflict largely idiosyncratic or perhaps path-dependent, in which case our efforts may be fruitless? In short, are we trying to forecast the behavior of “clouds” – “highly irregular, disorderly, and more or less unpredictable” phenomena – or the one of clocks – “regular, orderly, and highly predictable in their behaviour” (Popper [37])?

I argue in this essay that while efforts at forecasting should remain at the core of research on conflict, more work needs to be devoted to understanding the limits of conflict predictability. I first argue that too much effort has been devoted to explaining conflict, to the detriment of forecasting, even though both are necessary to understanding conflict. I then review some of the main approaches to predicting conflict, together with their respective limitations. Finally, I discuss the possibility that there might be inherent limitations to our ability to forecast conflict, and propose strategies to uncover these limitations.

## 1. To explain or to predict conflict?

The literature on conflict has largely focused on explaining the causes and course of conflict, rather than on attempting to predict them.<sup>1</sup> In particular, most of the work on conflict – whether it be the onset, course, or termination of civil or interstate wars – has been concerned with causal inference. Researchers start with a theory about the relationship between two variables such as military spending, trade patterns, or alliances, and then look for empirical associations between them as evidence of a causal mechanism – typically using various regression techniques (e.g. Bennett and Stam [3]). The role of the theory is central because of the high risk of spurious relationships. Indeed, human activities are strongly coupled and measurement errors are often inherent to the data, such that variables in the social sciences are particularly prone to covarying despite the absence of any causal link. Without a strong underpinning theory, then, the correlation observed between ( $X$ ) and ( $Y$ ) may simply be coincidental.

Yet the emphasis on theory and causal explanation has been excessive. Prediction has typically been considered either as unscientific (e.g. Shmueli [48]), or as pointless because of the perceived impossibility of forecasting political events (see Puchala [38], and for other examples: Ray and Russett [39] and Schneider, Gleditsch, and Carey [42]). This focus on theoretically-motivated approaches could be justified if, as is typically assumed, statistical models with high explanatory power also had high predictive power. Yet this is often not the case, and relying on p-values for the purposes of policy formulation is risky at best. In fact, some of the most prominent studies on conflict have poor out-of-sample forecasts, despite their strong theoretical backing and empirical evidence (Ward, Greenhill, and Bakke [60]).

This neglect of prediction is also detrimental to scientific progress, even from an explanatory standpoint. Both explanation and prediction are needed to generate and test theories (Ray and Russett [39]; Shmueli [48], p. 289).<sup>2</sup> In particular, predictions help to compare competing theories, whereas explanatory models can be difficult to compare, confirm, or contradict. In technology and medicine, theory is often derived from practice, and not the other way around (Taleb [54]). In the same way, new theories may well come from competing approaches to forecasting conflict. The test of accurate and reliable predictions also prevents the temptation to overfit past data at the cost of external validity on unobserved sets. In that sense, forecasting performances serve as Occam’s razor to select theories.

Prediction and explanation are two important dimensions of scientific inquiry that are not always perfectly aligned, and hence that should both be reported. First, certain models may offer little to no explanatory power, yet provide accurate forecasts (e.g. neural networks, as in Schrodt [45] or Beck, King, and Zeng [2]). This is, however, not a guarantee that they are the “right” model. Indeed, the

---

<sup>1</sup>Explanation is concerned about the (usually causal) effect of variables ( $X$ ) on an outcome variable ( $Y$ ). Prediction, on the other hand, is concerned about estimating a variable ( $Y$ ) for new observations. Typically this involves estimating a model from a learning set, and using the estimated model to form predictions about the future or new data (the testing set).

<sup>2</sup>In fact, Akaike even thought that the goal of statistics should not be to infer the true model – i.e., it should not be description of existing data – but to predict future data as accurately as possible.

‘wrong’ model can actually yield better predictions than the one closer to the underlying data generating process (Shmueli [48]).

On the other hand, models that explain but cannot predict can be useful. There are two main reasons why a model might explain but fail to predict. First, the information needed for prediction may simply be impossible to obtain. In international relations, for example, information about private knowledge and its extent is, almost by definition, impossible to obtain. As a result, some theories of signaling and private information are nearly impossible to test, yet provide valuable frameworks to think about threats and dispute escalation. Another reason why theories might explain without predicting is the idiosyncrasy of the phenomenon. The personality of a leader, for example, can hardly be predicted, yet a given personality may well explain a set of outcomes.

In short, explanation and prediction are two valuable yet distinct sides of scientific inquiry on the causes and course of conflict. Unfortunately, good forecasting papers are often rejected for their lack of a strong theory, yet theory-based papers rarely include out of sample tests. This double standard is detrimental to scientific progress, and I now report some of the most common approaches used in conflict forecasting.

## 2. Approaches to conflict forecasting

### 2.1. Experts

To the public, the most visible forecasters of political events are experts or pundits, whose opinions and predictions are typically relayed in the media. While they often provide valuable insights into a region or the relationship between two countries, there is no evidence that these insights have any predictive power. In fact, a study of predictions by 300 experts in politics and economics over the course of more than two decades revealed that the predictive performance of these experts on geopolitical events was on average no better than random guesses (Tetlock [55]). Experts who were successful in one period typically reverted to the mean in the next. Overall they performed no better than novices, and simple algorithms easily out-predicted them – a finding confirmed specifically in the context of conflict forecasting (Green and Armstrong [25]).

It is in fact not surprising that experts are poor predictors. First, they have an incentive to adopt a strong position, or else risk making dull predictions that are unlikely to receive media attention. Second, they rarely suffer the consequences of their failures. Correct forecasts are often hailed in the media, but wrong ones are promptly forgotten.

This is not to say, however, that experts, and more generally human forecasters, are useless. First, they may be more likely to offer dissonant opinions that ultimately contribute to forging the correct forecasts. The experts’ opinions, in other words, despite their poor predictive value, may contribute to a better performance by algorithms or ‘superforecasters’. But perhaps more importantly, a careful iterative selection of human predictors, for example via a tournament, has shown that they can produce consistently better forecasts than a wide range of competitors – including some of the most sophisticated algorithms (Tetlock and Gardner [56]).

### 2.2. Econometric approaches

Scholars and policy-makers have also relied on algorithms analyzing large amounts of data using various techniques, ranging from the more traditional regression techniques – typically logistic, due to

the binary nature of the outcome – to more intricate random forest or neural network models (Rummel [41]; Brandt and Freeman [6]; Muchlinski et al. [34]; Hegre et al. [28]). In the fields of international conflict and civil war, the majority of this work has until recently relied on structural variables such as regime type, GDP, ethnicity, or terrain (Beck, King, and Zeng [2]; O’Brien [35]). Unfortunately, these variables are typically measured yearly and vary slowly or not at all (Gleditsch and Ward [20]). As a result, while they can identify the situations that are conflict-prone, they generally fail to pinpoint which of them will actually break into war, and when (Chadeaux [14]).

Scholars have therefore increasingly relied on event data disaggregated at the temporal and spatial levels.<sup>3</sup> This initially involved the human coding of large corpuses of journalistic and historical documents into various categories of international events. Advances in computational methods and in particular in text parsing have made it possible to automate these processes, and hence to analyze far larger and broader ‘live’ sources of information, and thereby to move from structural to short-term measures of tensions and other markers of conflict. **TABARI** (Textual Analysis by Augmented Replacement Instructions, Schrodtt [44]), for example, uses the lead sentence of wire service reports (e.g. Reuters, Agence France press, etc.) to generate such event data. The **World-Wide Integrated Crisis Early Warning System (ICEWS)** is currently the most prominent of these event datasets (Boschee et al. [5]). Sponsored by the Defense Advanced Research Projects Agency in the United States, it provides a detailed database of political events at the sub-daily and subnational level (O’Brien [36]).

Approaches based on these datasets, and more generally on news-based sources, have scored some predictive successes (e.g. Schrodtt and Gerner [46]; Weidmann and Ward [63]; Brandt, Freeman, and Schrodtt [7]; Ward et al. [61]; Chadeaux [12]; Chiba and Gleditsch [15]). Yet their automatic approach to coding sources faces difficulties as well. In particular, scholars tend to rely on secondary sources such as news reports, which are neither comprehensive nor unbiased. Moreover, while temporal disaggregation has been relatively successful (there is generally little ambiguity about the timing of an attack or a speech), pinpointing the location of that event has proved to be much more challenging and error-prone (Weidmann [62]; Lee, Liu, and Ward [32]). Machine coding also still cannot analyze sentences with a complex structure, and typically ignores the connection between them.

Information derived from social media has also received increasing attention. Its advantage is obvious: the amount of data is considerable and transforms every user into a potential reporter – widely expanding the potential range of events covered. The amount of time that elapses between the event and the reporting is also greatly reduced, making real-time predictions possible in principle – an appealing proposition for policy-makers. Yet the challenges are also formidable. Most of what is written on social media is either irrelevant for the purposes of conflict forecasting – tweets, for example, are massively polluted by spam – or is deliberately false information planted by the government for purposes of manipulation (Spaiser et al. [52]).

### 2.3. Modeling war

Another approach to forecasting conflict relies on various forms of modeling. Game theory, in particular, has proven to be a core tool in the forecaster’s toolbox. An important advantage of game theory is that it forces the forecaster to model the interaction in great detail, and avoid omitting any relevant detail. Typically this involves two distinct processes. First, expert knowledge on a region or country is

---

<sup>3</sup>For a good review of available conflict data, see [http://www.nber.org/ens/feldstein/ENSA\\_Dataset.pdf](http://www.nber.org/ens/feldstein/ENSA_Dataset.pdf).

obtained to identify the relevant set of actors, together with their preferences and power. These elements are then incorporated in a model that simulates the interactions between these actors – for example a bargaining or coalition model. The model then returns a prediction about the likely outcome – whether it be the result of an election or a policy choice.

This approach has claimed a large number of successes (Gurr and Lichbach [26]), in particular when it comes to forecasting single events (Bueno de Mesquita [9]). Game theory software, for example, contributed to finding Osama bin Laden’s hiding location in Pakistan.<sup>4</sup> Yet, the overall evidence is mixed (Green [23]; Goodwin [22]; Green [24]). Part of the reason may be that the assumptions underlying game theory are at odds with empirical observations of human behavior, as behavioral economics has made clear (Kahneman [30]). For example, game theory may not be “able to trap the subtleties of [...] an international treaty bargaining session” (Shubik [49], p. 14).

Game theoretic models are also often limited by their complexity. They tend to require strong simplifying assumptions that limit their applicability and power. Agent-based models, which rely on simple rules but high computing power, can instead be used to model complex and emergent interactions. A drawback is that these models are typically so complex that our ability to draw inferences from them is limited. Yet if the point is to forecast, then agent-based models may be ideal in the sense that they allow us to run different scenarios and evaluate counterfactuals (Cederman [10]; Bhavnani et al. [4]).

#### 2.4. *Wise crowds?*

Of the many approaches to forecasting conflict, the most promising venues may be those that rely on their combination. In particular, the aggregation and weighting of various models and opinions appears to significantly improve predictions. Thus ensemble forecasts, which combine the predictions of various models, often outperform those of any model in isolation, and have been used with success in conflict prediction (Montgomery, Hollenbach, and Ward [33]).

Tournaments are another way of aggregating estimates. Tetlock and Gardner [56], for example, show that the recursive selection of human forecasters – repeatedly pitting the best of them against each other in iterative rounds – yields predictions that outperform not only experts, but also forecasts relying on algorithms and large databases.

Markets are also valuable tools to aggregate predictions, because they convert the wisdom of crowds into predictive probabilities. Financial markets are particularly good because they combine the forecasts of actors who have a financial stake in making accurate predictions. Securities are traded in a way that reflects the investors’ beliefs about the probability of a certain event occurring. Large events such as wars are economically and financially costly, and market participants will therefore strive to anticipate them as early as possible and to react accordingly. For example, bonds are likely to be sold in anticipation of a war (Chadeaux [13]), as will be the stocks of industries most likely to be affected by it (Schneider and Troeger [43]). Political prediction markets function in the same way as their financial peers, but can create securities that focus on questions of interest to political researchers (Arrow et al. [1]). For example, the Defense Advanced Research Projects Agency (DARPA) created the Policy Analysis Market to test the feasibility of an online market in which participants would bet on geopolitical trends or specific events.<sup>5</sup>

---

<sup>4</sup>See <http://www.economist.com/node/21527025>.

<sup>5</sup>See also the Iowa Electronic Markets (<http://tippie.uiowa.edu/iem/>), though it does not focus on conflict-related issues.

### 3. Can conflicts really be predicted?

While more and more research focuses on the prediction of inter- and intra-state conflicts, little attention has been paid to the potential limits of our ability to predict. Most work on forecasting takes the view that conflict is predictable, and that failures to forecast are due to our limited data and models. Yet, others argue that there may be inherent limitations to our ability to forecasts, and that our current successes might only be due to the fact that we are trying to predict the easy problems, neglecting the true challenges (Jäger [29]).<sup>6</sup>

#### 3.1. Are conflicts clocks, clouds, or swans?

The question of predictability ultimately hinges on the underlying nature of conflict. In the words of Popper, does it more closely resemble the world of clouds – “highly irregular, disorderly, and more or less unpredictable” – or the one of clocks – “regular, orderly, and highly predictable in their behaviour” (Popper [37]). Unfortunately we do not yet know which we are facing. On most days, international and domestic interactions resemble a clock. Small deviations are corrected in a reversion to the mean, and the stochastic process of daily events and tensions that may emerge on a local or global level is trend-stationary. Yet there are also rare shocks that do not follow this clock-like pattern. These are, of course, the events of interest here – conflicts, coups, acts of terrorism – that may start a cascade and change the clock into a cloud and the trend-stationary time series into unit-root processes (Doran [17]).

There is yet a third possibility: that conflicts are neither clouds nor clocks, but black swans (Taleb [53]). Black swans are game-changing events with such low probability that they cannot be predicted (even though experts often claim to have found obvious warning signs for them *ex post*). Black swans are different from simple rare events. While rare events occur infrequently, their probability is not low *conditional* on the relevant set of variables. On the other hand, black swans have a low probability even conditional on other variables.

Where conflict processes should be located on the clock-cloud-swan continuum matters. With clocks, predictions are possible, whether they be point or probabilistic predictions.<sup>7</sup> With clouds, the marginal cost of better predictions would be increasing, but we could at least learn about the aggregate distribution and data-generating process (e.g. Clauset, Young, and Gleditsch [16] on the frequency of terrorist events). With black swans, however, attempting to predict would be a fool’s game.

Several factors make it particularly challenging to predict conflict, and in fact may impose insuperable limits to our forecasts. First, our data are, almost by definition, prone to error and imprecision (e.g. Shellman, Hatfield, and Mills [47]). Part of it is due to poor measuring. But part of it is caused by strategic misrepresentation and concealing on the part of the relevant governments.

A second reason for the difficulty to predict conflicts is that their structure, and more generally the structure of international relations, are constantly evolving. The end of the cold war, for example, was largely unanticipated and challenged many of the structures and patterns that formed parts of the existing models. Even within a given conflict, the dynamic can dramatically change and necessitate a different model (e.g. the surge in Iraq, Bhavnani et al. [4]). The difficulty is that these changes are difficult to anticipate – they are often black swans themselves – such that our ability to forecast may be limited to the short-term. The long term, on the other hand, would be the result of too many compounding shocks to a point where predictions become futile. Ideally, our predictions would be able to accommodate these

<sup>6</sup>For related discussions, see Cederman and Weidmann [11] and Ward [59].

<sup>7</sup>See Brandt, Freeman, and Schrodt [8] on tools for density evaluations, as opposed to point predictions.

changes. This may require two-level predictions in which the structure itself is first predicted, and within that structure the short-term events would be forecasted with a different model. But of course this would compound the uncertainty about our model specification and data.

Another difficulty relates to the strategic nature of international relations and politics in general. First, actors are forward looking. They form their own predictions about the future, and act accordingly today. As a result, these predictions can affect their behavior today and invalidate these original predictions. As observers, then, we may have the right logic but end up not observing the phenomenon. Consider for example the problem of the onset of wars. If their contemporaries identify the underlying conditions as ripe for war, they may take additional steps to either prevent it, postpone its onset, or on the contrary speed it up, such that the initial predictions will be invalidated (e.g. Chadeaux [13] for empirical evidence of this pattern). The same logic also applies to the conduct and termination of war. The anticipations of forthcoming peace negotiations, for example, may lead certain actors – spoilers – to try harder to disrupt the peace process, thus reducing the prospects for peace (Kydd and Walter [31]).

Mixed strategies are another difficulty. States or domestic actors cannot always respond to the same situation in the same way, else their response becomes predictable and may be exploited by the adversary. Just like a tennisman will not always serve in the same place to prevent the opponent from anticipating his actions, leaders must vary their threats and responses to events. As a result, the same conditions and sequence of events may lead to different reactions, some potentially leading to wars whereas others do not. In such cases, probabilistic predictions remain possible, but point predictions are inherently impossible. While a large number of observations may bring us close to an estimate of the underlying probability distribution, the predictive value of our forecasts will be bound upward by a fundamental limit.

In fact, uncertainty itself may be necessary for the onset and continuation of war. Indeed, one of the central rationalist explanations for why bargaining might break down into war is incomplete information of at least one of the participants. As a result, “we cannot predict in individual cases whether states will go to war, because war is typically the consequence of variables that are unobservable *ex ante*, both to us as researchers and to the participants” (Gartzke [19], p. 567).

Conflicts and the processes leading to them may also be path-dependent. A small event may lead to a cascading effect and ultimately to war. Yet the same underlying structure could possibly have accommodated an alternate equilibrium in which peace prevailed. Self-reinforcing processes mean that international interactions may magnify the effects of chance. Looking back, we may be able to trace the explosion of a keg – conflict – to a single spark (though this is itself debatable, as evidenced by the unfaltering scholarship on the causes of WWI), but looking forward, we are unable to know which spark will ignite the keg. In the same way, seismologists understand the causes of earthquakes and are able to monitor seismological variations with high precision but still cannot predict their onset with much early warning.<sup>8</sup>

### 3.2. *How can we know?*

A core difficulty in determining whether we are dealing with clocks or black swans, and more generally in improving our forecasts, is that wars are rare events. Evaluating the predictive validity of a model may therefore require years or even decades. As a result, feedback is delayed and improvement slow. True long-term predictions such as Hegre et al. [27] (as opposed to forecasting exercises using

---

<sup>8</sup>Asset prices and bubbles in financial markets are also well-theorized, yet cannot be forecasted (Fama [18]; Sornette [51]).

cross-validation tests on past data) are thus particularly difficult to evaluate. However, new data disaggregated at the monthly and daily levels make it possible to test hypotheses and evaluate forecasts much more rapidly than before. Datasets such as ICEWS provide hundreds of events every day, such that forecasting competitions could be organized and evaluated within months.

The real difficulty, though, is not to know whether or not we predict well, but rather whether there are fundamental limits to our ability to predict. Knowing why we fail – because of poor models or the inherent impossibility to forecast – matters for several reasons. First, information about the limits of forecasting would help us decide which area of research is most promising and where to allocate our scarce research resources. Second, knowing these limits also matters because our scientific forecasts may in fact be harmful. Publishing forecasts of outcomes that are in fact unpredictable – and hence repeated failure to correctly predict – may lead to distrust of the public toward the scientific enterprise. Moreover, a good overall forecasting record may lead policy-makers to take into account the advice of the forecaster. Yet even a good overall forecasting record may mask systematically poor forecasts in a given area, due to the unpredictability of that area. If decision makers trust the overall record and follow the recommendation in that area, the results will potentially be disastrous. Knowing what is unpredictable would therefore avoid these costly mistakes.

We therefore need to study the limits of our own ability to know. At the theoretical level, we know that mixed strategies and incomplete information guarantee that the onset of war will never be perfectly predictable, as discussed above. Beyond theoretical arguments for the indeterminacy of international interactions, however, can we uncover empirically the limits of our ability to predict conflicts?

One approach is to simply evaluate the predictive performance of our best model, and assume that the errors are due to limits to the predictability of conflicts. This is the approach adopted by Song et al. [50], though in a different field. Most models forecasting the onset of conflict thus reach an area under the Receiver-Operating Characteristic (ROC) curve of approximately 80%, yet few manage to progress beyond that threshold. This may be the fundamental limit of predictability. The problem with this approach is that while failures to forecast may indeed be attributed to the inherent impossibility of forecasting conflict, nothing guarantees that they are not due to errors in our assumptions, models, or data.

A different approach consists in examining the marginal contribution of additional variables to predictive accuracy. From this, we can extrapolate the asymptotic limits of predictability from our current path of discovery. Chadeaux [13] thus shows that the size of marginal contribution of each additional variable is decreasing and quickly reaches zero, thus suggesting possible inherent limits to our capacity to predict the onset of conflict. But of course, there can never be any certainty that this path will not be upended by a new discovery, model, or technique.

Yet another method is to examine the past and evaluate the performance of past forecasters. Chadeaux [13], for example, shows that financial markets have underestimated the risk of conflict over the past 200 years, and that their forecasting ability has remained constant over time. This is particularly surprising, as financial market participants have a strong incentive to make correct forecast – and hence to improve on past performance. This inability to improve may be an indication of limitations inherent to the pattern leading to the onset of war, for example because of feed-forward effects.

Another approach consists in measuring our ability to forecast various conflict processes with different forecast lead times. If conflicts are cloud-like, then our ability to forecast should decrease exponentially as we make predictions farther out in time. If, on the other hand, they are clock-like, then we should expect the predictability to decline only linearly or as a function of the lead time.

#### 4. Conclusion

The field of international relations has long emphasized theory to the detriment of work on prediction. Fortunately, forecasts of conflicts are of increasing interest. Although most of that attention has come from practitioners in the defense and intelligence communities, academics have also forayed into this area with success.

Regardless of the method used (econometric, game theoretic, markets), an important advantage of the work on conflict forecasting is its cumulativeness – the ability to build upon existing models – and comparability – forecasts can be pitted against each other to select the best theory, method, or data.<sup>9</sup> In that regard, it is critical that data be made available publicly. It is especially important to pool together observations derived from our respective forecasts because forecasting rare events is slow to evaluate and necessitates a large number of observations.

To understand predictability, it is therefore important not only that purely predictive efforts be further encouraged, but also that negative results – the failure to improve upon existing forecasts – be reported. Only then will we be able to start inferring the bounds of our knowledge.

#### Acknowledgements

This research was supported by a grant from the Arts and Social Sciences Benefaction Fund at Trinity College Dublin.

#### References

- [1] K.J. Arrow, R. Forsythe, M. Gorham, R. Hahn, J.O. Hanson, R. Ledyard, S. Levmore, R. Litan et al., The promise of prediction markets, *Science* **320**(5878) (2008), 877. doi:[10.1126/science.1157679](https://doi.org/10.1126/science.1157679).
- [2] N. Beck, G. King and L. Zeng, Improving quantitative studies of international conflict: A conjecture, *American Political Science Review* **94**(1) (2000), 21–35. doi:[10.1017/S0003055400220078](https://doi.org/10.1017/S0003055400220078).
- [3] D.S. Bennett and A.C. Stam, *The Behavioral Origins of War*, University of Michigan Press, 2009. doi:[10.3998/mpub.23490](https://doi.org/10.3998/mpub.23490).
- [4] R. Bhavnani, K. Donnay, D. Miodownik, M. Mor and D. Helbing, Group segregation and urban violence, *American Journal of Political Science* **58**(1) (2014), 226–245. doi:[10.1111/ajps.12045](https://doi.org/10.1111/ajps.12045).
- [5] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz and M. Ward, *ICEWS Coded Event Data*, Harvard Dataverse, 2017. doi:[10.7910/DVN/28075](https://doi.org/10.7910/DVN/28075).
- [6] P.T. Brandt and J.R. Freeman, Advances in Bayesian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis, *Political Analysis* **14**(1) (2006), 1–36. doi:[10.1093/pan/mpi035](https://doi.org/10.1093/pan/mpi035).
- [7] P.T. Brandt, J.R. Freeman and P.A. Schrodtt, Real time, time series forecasting of inter- and intra-state political conflict, *Conflict Management and Peace Science* **28**(1) (2011), 41–64. doi:[10.1177/0738894210388125](https://doi.org/10.1177/0738894210388125).
- [8] P.T. Brandt, J.R. Freeman and P.A. Schrodtt, Evaluating forecasts of political conflict dynamics, *International Journal of Forecasting* **30**(4) (2014), 944–962. doi:[10.1016/j.ijforecast.2014.03.014](https://doi.org/10.1016/j.ijforecast.2014.03.014).
- [9] B. Bueno de Mesquita, *The Predictioneer's Game: Using the Logic of Brazen Self-Interest to See and Shape the Future*, Random House, New York, 2009. isbn:[9781400067879](https://www.isbn-international.org/product/9781400067879).
- [10] L.-E. Cederman, Endogenizing geopolitical boundaries with agent-based modeling, *Proceedings of the National Academy of Sciences* **99**(Suppl. 3) (2002), 7296–7303. doi:[10.1073/pnas.082081099](https://doi.org/10.1073/pnas.082081099).
- [11] L.-E. Cederman and N.B. Weidmann, Predicting armed conflict: Time to adjust our expectations?, *Science* **355**(6324) (2017), 474–476. doi:[10.1126/science.aal4483](https://doi.org/10.1126/science.aal4483).
- [12] T. Chadeaux, Early warning signals for war in the news, *Journal of Peace Research* **51**(1) (2014), 5–18. doi:[10.1177/0022343313507302](https://doi.org/10.1177/0022343313507302).

---

<sup>9</sup>In fact, these two aspects have contributed to the success of the [Good Judgement Project](#), in which forecasters select, compare, and improve their predictions over repeated rounds, in a way that leads to forecasts superior to those of any single model.

- [13] T. Chadeaux, Market anticipations of conflict onsets, *Journal of Peace Research* **54**(2) (2017), 313–327. doi:10.1177/0022343316687615.
- [14] T. Chadeaux, The Triggers of War: Disentangling the Spark from the Powder Keg, Working paper, Trinity College Dublin, 2015. Available at: <https://ssrn.com/abstract=2409005>.
- [15] D. Chiba and K. Gleditsch, The shape of things to come? Expanding the inequality and grievance model for civil war forecasts with event data, *Journal of Peace Research* **54**(2) (2017), 275–297. doi:10.1177/0022343316684192.
- [16] A. Clauset, M. Young and K.S. Gleditsch, On the frequency of severe terrorist events, *Journal of Conflict Resolution* **51**(1) (2007), 58–87. doi:10.1177/0022002706296157.
- [17] C.F. Doran, Why forecasts fail: The limits and potential of forecasting in international relations and economics, *International Studies Review* **1**(2) (1999), 11–41. doi:10.1111/1521-9488.00154.
- [18] E.F. Fama, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* **25**(2) (1970), 383–417. doi:10.2307/2325486.
- [19] E. Gartzke, War is in the error term, *International Organization* **53**(3) (1999), 567–587. doi:10.1162/002081899550995.
- [20] K.S. Gleditsch and M.D. Ward, Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes, *Journal of Peace Research* **50**(1) (2013), 17–31. doi:10.1177/0022343312449033.
- [21] J.A. Goldstone, R.H. Bates, D.L. Epstein, T.R. Gurr, M.B. Lustik, M.G. Marshall, J. Ulfelder and M. Woodward, A global model for forecasting political instability, *American Journal of Political Science* **54**(1) (2010), 190–208. doi:10.1111/j.1540-5907.2009.00426.x.
- [22] P. Goodwin, Forecasting games: Can game theory win?, *International Journal of Forecasting* **18**(3) (2002), 369–374. doi:10.1016/S0169-2070(02)00022-5.
- [23] K.C. Green, Forecasting decisions in conflict situations: A comparison of game theory, role-playing, and unaided judgement, *International Journal of Forecasting* **18**(3) (2002), 321–344. doi:10.1016/S0169-2070(02)00025-0.
- [24] K.C. Green, Game theory, simulated interaction, and unaided judgement for forecasting decisions in conflicts: Further evidence, *International Journal of Forecasting* **21**(3) (2005), 463–472. doi:10.1016/j.ijforecast.2005.02.006.
- [25] K.C. Green and J.S. Armstrong, The ombudsman: Value of expertise for forecasting decisions in conflicts, *Interfaces* **37**(3) (2007), 287–299. doi:10.1287/inte.1060.0262.
- [26] T.R. Gurr and M.I. Lichbach, Forecasting internal conflict: A competitive evaluation of empirical theories, *Comparative Political Studies* **19**(1) (1986), 3–38. doi:10.1177/0010414086019001001.
- [27] H. Hegre, H. Buhaug, K.V. Calvin, J. Nordkvelle, S.T. Waldhoff and E. Gilmore, Forecasting civil conflict along the shared socioeconomic pathways, *Environmental Research Letters* **11**(5) (2016). doi:10.1088/1748-9326/11/5/054002.
- [28] H. Hegre, N.W. Metternich, H.M. Nygård and J. Wucherpfennig, Introduction: Forecasting in peace research, *Journal of Peace Research* **54**(2) (2017), 113–124. doi:10.1177/0022343317691330.
- [29] K. Jäger, Not a new gold standard: Even big data cannot predict the future, *Critical Review* (2016), 1–21. doi:10.1080/08913811.2016.1237704.
- [30] D. Kahneman, *Thinking, Fast and Slow*, Macmillan, 2011. isbn:9780374533557.
- [31] A. Kydd and B.F. Walter, Sabotaging the peace: The politics of extremist violence, *International Organization* **56**(02) (2002), 263–296. doi:10.1162/002081802320005487.
- [32] S.J. Lee, H. Liu and M.D. Ward, Lost in Space: Geolocation in Event Data, 2016, CoRR. Available at: <http://arxiv.org/abs/1611.04837>.
- [33] J.M. Montgomery, F.M. Hollenbach and M.D. Ward, Improving predictions using ensemble Bayesian model averaging, *Political Analysis* **20**(3) (2012), 271–291. doi:10.1093/pan/mps002.
- [34] D. Muchlinski, D. Siroky, J. He and M. Kocher, Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data, *Political Analysis* **24**(1) (2016), 87–103. doi:10.1093/pan/mpv024.
- [35] S.P. O'Brien, Anticipating the good, the bad, and the ugly: An early warning approach to conflict and instability analysis, *Journal of Conflict Resolution* **46**(6) (2002), 791–811. doi:10.1177/002200202237929.
- [36] S.P. O'Brien, Crisis early warning and decision support: Contemporary approaches and thoughts on future research, *International Studies Review* **12**(1) (2010), 87–104. doi:10.1111/j.1468-2486.2009.00914.x.
- [37] K.R. Popper, *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, 1972. isbn:0198243707.
- [38] D.J. Puchala, Woe to the orphans of the scientific revolution, *Journal of International Affairs* (1990), 59–80, <http://www.jstor.org/stable/24357224>.
- [39] J.L. Ray and B. Russett, The future as arbiter of theoretical controversies: Predictions, explanations and the end of the cold war, *British Journal of Political Science* **26**(4) (1996), 441–470. doi:10.1017/S0007123400007560.
- [40] N. Rost, Will it happen again? On the possibility of forecasting the risk of genocide, *Journal of Genocide Research* **15**(1) (2013), 41–67. doi:10.1080/14623528.2012.759398.
- [41] R.J. Rummel, Forecasting international relations: A proposed investigation of three-mode factor analysis, *Technological Forecasting* **1**(2) (1969), 197–216. doi:10.1016/0099-3964(69)90020-9.
- [42] G. Schneider, N.P. Gleditsch and S. Carey, Forecasting in international relations: One quest, three approaches, *Conflict Management and Peace Science* **28**(1) (2011), 5–14. doi:10.1177/0738894210388079.

- [43] G. Schneider and V.E. Troeger, War and the world economy stock market reactions to international conflicts, *Journal of Conflict Resolution* **50**(5) (2006), 623–645. doi:[10.1177/0022002706290430](https://doi.org/10.1177/0022002706290430).
- [44] P. Schrodt, TABARI: Textual Analysis by Augmented Replacement Instructions, 2009, <http://eventdata.parusanalytics.com/tabari.dir/TABARI.0.8.4b3.manual.pdf>.
- [45] P.A. Schrodt, Prediction of interstate conflict outcomes using a neural network, *Social Science Computer Review* **9**(3) (1991), 359–380. doi:[10.1177/089443939100900302](https://doi.org/10.1177/089443939100900302).
- [46] P.A. Schrodt and D.J. Gerner, Cluster-based early warning indicators for political change in the contemporary levant, *American Political Science Review* **94**(4) (2000), 803–817. doi:[10.2307/2586209](https://doi.org/10.2307/2586209).
- [47] S.M. Shellman, C. Hatfield and M.J. Mills, Disaggregating actors in intranational conflict, *Journal of Peace Research* **47**(1) (2010), 83–90. doi:[10.1177/0022343309350029](https://doi.org/10.1177/0022343309350029).
- [48] G. Shmueli, To explain or to predict?, *Statistical Science* **25**(3) (2010), 289–310. doi:[10.1214/10-STS330](https://doi.org/10.1214/10-STS330).
- [49] M. Shubik, *Games for Society, Business, and War*, Elsevier, 1975. isbn:[9780444412850](https://www.isbn-international.org/details/9780444412850).
- [50] C. Song, Z. Qu, N. Blumm and A.-L. Barabási, Limits of predictability in human mobility, *Science* **327**(5968) (2010), 1018–1021. doi:[10.1126/science.1177170](https://doi.org/10.1126/science.1177170).
- [51] D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press, 2009. isbn:[9781400829552](https://www.isbn-international.org/details/9781400829552).
- [52] V. Spaier, T. Chadeaux, K. Donnay, F. Russmann and D. Helbing, Communication power struggles on social media: A case study of the 2011-12 Russian protests, *Journal of Information Technology & Politics* (2017), <https://ssrn.com/abstract=2528102>.
- [53] N.N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, New York, 2011. isbn:[9780812973815](https://www.isbn-international.org/details/9780812973815).
- [54] N.N. Taleb, *Antifragile: Things That Gain from Disorder*, Random House, 2012. isbn:[9780812979688](https://www.isbn-international.org/details/9780812979688).
- [55] P.E. Tetlock, *Expert Political Judgment*, Princeton University Press, Princeton, NJ, 2005. isbn:[9780691128719](https://www.isbn-international.org/details/9780691128719).
- [56] P.E. Tetlock and D. Gardner, *Superforecasting: The Art and Science of Prediction*, Random House, 2016. isbn:[9780804136716](https://www.isbn-international.org/details/9780804136716).
- [57] J. Ulfelder, Forecasting Onsets of Mass Killing, Working paper, 2012. Available at: <https://ssrn.com/abstract=2409005>. doi:[10.1371/journal.pone.0154148](https://doi.org/10.1371/journal.pone.0154148).
- [58] T. Van Holt, J.C. Johnson, S. Moates and K.M. Carley, The role of datasets on scientific influence within conflict research, *PloS One* **11**(4) (2016).
- [59] M.D. Ward, Can we predict politics? Toward what end?, *Journal of Global Security Studies* **1**(1) (2016), 80–91. doi:[10.1093/jogss/ogv002](https://doi.org/10.1093/jogss/ogv002).
- [60] M.D. Ward, B.D. Greenhill and K.M. Bakke, The perils of policy by *P*-value: Predicting civil conflicts, *Journal of Peace Research* **47**(4) (2010), 363–375. doi:[10.1177/0022343309356491](https://doi.org/10.1177/0022343309356491).
- [61] M.D. Ward, N.W. Metternich, C.L. Dorff, M. Gallop, F.M. Hollenbach, A. Schultz and S. Weschle, Learning from the past and stepping into the future: Toward a new generation of conflict prediction, *International Studies Review* **15**(4) (2013), 473–490. doi:[10.1111/misr.12072](https://doi.org/10.1111/misr.12072).
- [62] N.B. Weidmann, On the accuracy of media-based conflict event data, *Journal of Conflict Resolution* **59**(6) (2015), 1129–1149. doi:[10.1177/0022002714530431](https://doi.org/10.1177/0022002714530431).
- [63] N.B. Weidmann and M.D. Ward, Predicting conflict in space and time, *Journal of Conflict Resolution* **54**(6) (2010), 883–901. doi:[10.1177/0022002710371669](https://doi.org/10.1177/0022002710371669).