Toward a Bayesian Network Model of Events in International Relations

Ali Jalal-Kamali and David V. Pynadath

Institute for Creative Technologies, University of Southern California
{jalalkam,pynadath}@usc.edu

Abstract. Formal models of international relations have a long history of exploiting representations and algorithms from artificial intelligence. As more news sources move online, there is an increasing wealth of data that can inform the creation of such models. The Global Database of Events, Language, and Tone (GDELT) extracts events from news articles from around the world, where the events represent actions taken by geopolitical actors, reflecting the actors' relationships. We can apply existing machine-learning algorithms to automatically construct a Bayesian network that represents the distribution over the actions between actors. Such a network model allows us to analyze the interdependencies among events and generate the relative likelihoods of different events. By examining the accuracy of the learned network over different years and different actor pairs, we are able to identify aspects of international relations from a data-driven approach. We are also able to identify weaknesses in the model that suggest needs for additional domain knowledge.

Keywords: Intelligent Agents, Bayesian Networks, Modeling and Simulation, International Relations

1 Introduction

Formal models of international relations have a long history of exploiting representations and algorithms from artificial intelligence [1, 3, 9, 11]. For example, game-theoretic models have supported prescriptive analyses of foreign-policy decisions (e.g., [10]). Political scientists have also used rule-based systems to build descriptive models of the behaviors of geopolitical actors [3]. These manually created models demonstrate the value that AI methodologies can provide in the study of international relations.

As more and more news sources move online, there is increasing data that can inform the creation of such models. More importantly, these data are now in computer-readable formats that can potentially support the *automatic* creation of models. For example, the Global Database of Events, Language, and Tone (GDELT) and the iData repository of Integrated Crisis Early Warning System (ICEWS) both represent hundreds of millions of actions in over 300 categories (e.g., negotiation, accusations, military deployment) taken by geopolitical actors (countries, international organizations, etc.), often directed at other such actors.[1]

[1] gdeltproject.org, lockheedmartin.com/us/products/W-ICEWS/iData.html

A computational model of the likelihoods of different action types would be invaluable in describing the behavior of actors in international relations. For example, one might expect two allied nations to be more likely to engage in trade agreements and other cooperative actions, as opposed to nations with a more hostile relationship. The dependency may operate in the opposite direction, too, where the relationship between two nations is likely to suffer if one nation makes an accusatory statement about the other. Representing this complex interdependence among the types of actions and the actors' relationship is critical to building an accurate model.

In this work, we apply existing algorithms to learn a Bayesian network [4–6] that represents the distribution over the actions between geopolitical actors. Such a network allows us to analyze the interdependencies among events and generate relative likelihoods of different event types between two actors. We focus on our use of GDELT as a source of events that we translate into a categorical distribution of the actions two actors take toward each other[2]. Besides the events extracted by GDELT, our network model also represents the Ideal Point distances, a measure of country affinity that researchers in international relations derive from UN voting records.

By examining the accuracy of the learned network over different years and different pairs of actors, we are able to identify aspects of international relations that are quantifiable from a purely data-driven approach. By leveraging the declarative nature of our Bayesian network model, we are able to inspect its dependency structure and draw conclusions that provide insight into the types of events that are most strongly tied to geopolitical relationships. Furthermore, we can exploit existing Bayesian network inference algorithms to compute any conditional probability of interest, allowing us to analyze the joint distribution of events to gain insight into various event dependencies. We are also able to identify weaknesses in the model that suggest needs for additional domain knowledge. This work thus represents an important first step toward making use of advances in AI methods in the modeling of international relations.

2 International Relations Data

The "Global Database on Events, Location, and Tone" (GDELT) contains international events automatically extracted on a daily basis from different news sources around the world, dating back to 1979. We restrict our investigation to pairwise international relations, so we use only events that list two countries as the actors. Each event's type is in the form of a numeric code, categorizing the action according to the Conflict and Mediation Event Observations (CAMEO) framework[3]. The CAMEO event codes constitute a hierarchy, ranging from top-level categories like "Appeal" (02) and "Fight" (19), to intermediate categories like "Appeal for material cooperation" (021) and "Occupy territory" (192). The CAMEO framework contains 310 such categories.

For each year of GDELT events, we aggregate the events for each pair of actors (ignoring their order in this investigation). For each such pair of actors, we compute a histogram of the number of occurrences of each event type and then normalize these event counts to be a percentage. This normalization loses the potential information contained

[2] We also used ICEWS, but omit those results for space considerations.

[3] http://eventdata.parusanalytics.com/data.dir/cameo.html

in the volume of events between actors (e.g., a qualitative difference in the relationship between the USA and the UK vs. between the USA and Turkmenistan based on only the number of actions). On the other hand, the normalization allows us to potentially generalize across relationships that are similar in character, if not in volume.

We also include *Ideal Point* distances, a measure of affinity that political scientists derive from applying a distance metric to voting in the General Assembly of the United Nations.[4] Our goal is a country-independent model of relative event likelihoods in combination with these Ideal Point distances. We collect data across pairwise relationships between countries from the years 2006–2012, which have comparable quantities of annual GDELT data, while also having available Ideal Point data. Although there are 310 different categories in the CAMEO event code list, many categories are not represented in our target data sets. To ensure consistency of representation across these years, we selected the CAMEO event codes that have appeared at least once in all of the years from 2006–2012, leaving an intersection of 133 event types. After including the Ideal Point distance, we arrive at 134 variables for each pair of countries.

To avoid distortions due to country pairs for which GDELT has very few (i.e., non-representative) events, we use a minimum threshold for actions between countries, considering both the total count of the events (to avoid small samples) and the percentage of non-zero events (to avoid highly skewed histograms) for each pair of countries. After examining the size of the data sets resulting from alternate threshold settings, we arrived at 50 for the minimum total count of events, and 5% as the minimum percentage of the event categories present in a country relationship for every year. The resulting numbers of country relationships for 2006–2012 are respectively: 773, 886, 1014, 1316, 1198, 1300, 1310. Table 1 shows a subset of the data for 2006.

Country Pairs		02: Appeal	021: Appeal for material cooperation	...	Ideal Point Distance
Afghanistan	China	0.017	0.027	...	0.167
Argentina	Australia	0.033	0.010	...	0.788
...	...	...	...	...	...
Yemen	Qatar	0.018	0.000	...	0.279

Table 1. Sample data entries from 2006.

3 Learning a Bayesian Network Model

To study the causality and dependency structure within the event categories and Ideal Point distances, we seek a model in the form of a Bayesian Network [4–6]. A Bayesian network provides a compact graphical representation of a joint probability distribution over a set of random variables. By representing such a distribution over our international relations variables, we can use standard algorithms to answer queries about the conditional probability of event categories or UN voting patterns of interest given the occurrence of other event categories. For example, we may use such a network to examine the potential likelihood of cooperation vs. conflict, contingent on the frequency of appeals, public statements, and other actions in the recent history between two countries. Unlike a classifier approach to the problem, the Bayesian network representation allows us to interchangeably treat any variable as the input or output to our queries.

[4] https://dataverse.harvard.edu/dataverse/Voeten

In addition to performing such probabilistic inference, the Bayesian network's graphical structure can itself provide insight into the underlying process. In particular, the directed edges in the Bayesian network reflect properties of conditional independence among the variables. By studying the link structure of the network, we can get a better understanding of the causal process that generates the distribution being modeled [7]. Therefore, if we can construct an accurate Bayesian network model of the distribution over event categories and Ideal Point distances, then the resulting graph structure may reveal interesting properties underlying the behavior selection of geopolitical actors.

Another advantage of Bayesian networks is the development of algorithms that can automatically learn the best structure to represent a data set [2, 5]. In this work, we use the algorithms contained in the *bnlearn* R package (www.bnlearn.com). All of our variables are continuous, so we approximate the distributions over them as Gaussians. In terms of the specific *bnlearn* algorithms, we treat all of the variables as having linear conditional Gaussian distributions, where the mean of each child is a linear function of its parents' values. While this assumption is likely to be overly strong, it provides a good first approximation. For each year's training data, after learning the network structure model, we fit the model to the data to obtain the standard deviations, intercepts, and the coefficients of each node's parents to evaluate the means. We can then measure the probability of a different year's test data given the learned network.

4 Accuracy of Bayesian Network Models

While a Bayesian network representation is capable of capturing the interdependence among the relative frequencies of different action categories, that capability will not be useful if we do not have data that supports the learning of such dependencies. To quantify the ability of the learning algorithm to capture this interdependence, we compare the performance of a Bayesian network model against a model that assumes independence among the variables. We can view the latter as learning a Bayesian network with no links among the variables. We hope that the Bayesian network with a learned link structure will provide a better explanation of our test data than one without any links.

To quantify how well a given model explains our test data, we compute a joint probability of the data set. For continuous-valued variables, we integrate the appropriate Probability Density Function (PDF) over the 1%-wide interval in which the observed value falls. Table 2 presents the log of the mean probability of a given year's test data with respect to the models (independence in Column 2, Bayesian network in Column 3) learned from the other years' training data. The Bayesian network model consistently outperforms the independence model by multiple orders of magnitude, providing strong evidence that the dependency structure provides critical information.

$\log(\text{mean}(\Pr(\text{test}\mid\text{train}))$			
Year	**Independence**	**BN (Same)**	**BN (Different)**
2006	-62.09	-59.26	-61.80
2007	-64.82	-62.21	-60.85
2008	-66.00	-62.87	-60.42
2009	-63.34	-59.94	-60.20
2010	-63.55	-60.73	-60.56
2011	-62.40	-59.64	-60.26
2012	-63.09	-60.68	-60.70

Table 2. Accuracy of learned models.

5 Analysis of Bayesian Network Models

While our Bayesian networks offer predictive value, there is still room for improvement in their accuracy. However, the very significant improvement gained by the networks' dependency structure provides evidence that we can extract insight by inspecting the networks themselves. Furthermore, such examination can also inform us as to where our network models are doing well and where they are doing poorly.

5.1 Variations in Models over Time

Ideally, we would arrive at a time-invariant model, so that we can reuse the same model year after year, simply by providing it with the given year's event data. Given this goal, one question that arises is the degree to which the Bayesian network models are generalizable over time. To answer this question, we first took each of the seven network and independence models from Table 2 and measured their accuracy against the test data from *other* years as well. For each year of test data, the Bayesian networks outperform the best independence model by multiple orders of magnitude. In fact, the accuracy rank of each year's model, both for the Bayesian network and independence models, is consistent across test sets, an interesting phenomenon to study in future work.

First Event		**Second Event**
Host a visit	$\rightarrow$	Make a visit
Use conventional military force	$\rightarrow$	Express intent to meet or negotiate
Use conventional military force	$\rightarrow$	Fight with small arms and light weapons
Abduct hijack or take hostage	$\leftrightarrow$	Return release not specified below
Accuse	$\leftrightarrow$	Make a visit
Allow international involvement	$\leftrightarrow$	Provide military protection
Allow international involvement	$\leftrightarrow$	Provide military aid
Arrest/detain/charge w/legal action	$\leftrightarrow$	Consult
Arrest/detain/charge w/legal action	$\leftrightarrow$	Engage in negotiation
Arrest/detain/charge w/legal action	$\leftrightarrow$	Express intent to meet or negotiate
Arrest/detain/charge w/legal action	$\leftrightarrow$	Make a visit
Consult	$\leftrightarrow$	Criticize or denounce
Consult	$\leftrightarrow$	Engage in negotiation
Consult	$\leftrightarrow$	Return release
Consult	$\leftrightarrow$	Use conventional military force
Consult not specified below	$\leftrightarrow$	Use conventional military force
Cooperate economically	$\leftrightarrow$	Express intent to engage in material cooperation
Express intent to cooperate	$\leftrightarrow$	Sign formal agreement
Engage in negotiation	$\leftrightarrow$	Make a visit
Make a visit	$\leftrightarrow$	Use conventional military force

Table 3. Common links in models over years 2006-2012.

We can also investigate the generalizability of the network structures by learning a Bayesian network for a given year's data (e.g., 2006), and then using the resulting structure to constrain the learning of a network for a different year (e.g., 2007). The learning of the second network can thus modify the parameters on the links, but not the link structure itself. We then evaluate this second Bayesian network on a third year's data (e.g., 2008). Table 2's fourth column presents the average probabilities over the test sets for each structure- and parameter-learning pair. Even when imposing a different year's learned structure, the resulting networks still outperform the independence model. In fact, the networks learned when using different years for the structure- and parameter-learning show much less variance than and sometimes outperform the original networks. This result show more encouraging evidence of time-invariant properties of the link structure, although further investigation is necessary.

We can also directly examine the links to see which dependencies are consistently present across the set of networks. Table 3 lists the 20 (out of a possible 8911) links that exist in each of the 7 networks, of which only the first 3 occur with the same direction. The Bayesian network structure thus provides us with potential insight into event types and interdependencies that are exhibited most frequently in the GDELT data set. It is important to note that BN links are a subset of dependencies, so non-BN methods cannot arrive at the same results. In particular, the absence of links does not represent independence, but rather conditional independence. So the BN algorithm in a way finds the most direct influences, or causal influences.

5.2 Variations in Models over Different Countries

We seek a model that is not just time-invariant, but also actor-independent. In this section, we investigate whether there are certain actor relationships for which our models perform better than others. Table 4 shows a partial ranking of the relationships whose event histograms are given the highest and lowest probabilities by our learned networks for 2006. Table 4's event counts suggests that the highest-ranked pairs performed many more actions than the lowest-ranked pairs, even though our input data contains no information about the volume of actions between actors due to normalization. More precisely, the correlation between the number of actions between two actors and the pair's rank in our model's accuracy ranges over $[-0.26, -0.11]$ over the different years of data. In other words, our model more accurately predicts the action breakdown between actors for which we have more events in GDELT. This correlation is encouraging in that it suggests that a significant part of the inaccuracy of our model derives from actor relationships from which we have limited observations. In other words, we might expect our model to perform better if we were able to get a more accurate categorization of their actions by GDELT, since the algorithms used in GDELT determine the accuracy of the event categorization.

Relationship	**Rank**	**Count**
USA-CAN	1	8,119
USA-UKG	2	21,176
IRN-AFG	3	2,704
USA-RUS	4	20,336
RUS-BLR	5	3,482
...	...	...
CAN-FIN	763	57
CAN-CHL	764	60
ISR-FSM	765	75
BEL-SEN	766	80
CAN-CUB	767	55

Table 4. Ranking of relationships by accuracy in 2006.

5.3 Ideal-Point-Distance Dependencies

The Bayesian network structure also allows us to look at the dependencies of specific nodes of interest. For example, we can inspect the Markov blanket of the Ideal Point distance node, i.e., its parents, children, and immediate parents of those children. Across the 2006–2012 models, there are a total of 129 events that have appeared in the Markov blanket of the Ideal Point distance variable. Table 5 contains the events that have appeared in the Markov blanket of the Ideal Point distance node at least five times. The size of the Markov blanket ranges from 10–128 nodes over all of the years, but if we ignore the 2008 network, the range narrows to 10–37, representing a much smaller subset than the 133 overall event types. There is one variable, corresponding to the event category "Make optimistic comment", that appears in the Markov blanket in all of the learned networks. There is obviously some consistency across these networks in terms of which nodes are connected to the Ideal point distance node. This consistency suggests that there is some more general dependency between actions of these identified categories and the UN voting patterns measured by the Ideal Point methodology. This dependency suggests an interesting line of investigation that can be informed by political science theories underlying that methodology.

Event Type	Count	Impact
Make optimistic comment	7	.0710
Meet at a third location	6	-.0237
Sign formal agreement	6	-.0592
Criticize or denounce	5	.0626
Fight with artillery and tanks	5	.0353
Provide aid	5	.0353
Make statement	5	.0327
Impose embargo, boycott, or sanctions	5	.0280
Use conventional military force	5	.0248
Demand	5	.0216
Employ aerial weapons	5	.0175
Reduce or break diplomatic relations	5	.0160

Table 5. Markov blanket of ideal points.

The Markov blanket also provides a sufficient subnetwork for the Ideal Point distance node, which is conditionally independent of all other nodes in the network given the variables in its Markov blanket [8]. To see how effective this subnetwork is in predicting the existing Ideal point distances from GDELT events, we learned a Bayesian network over the aggregation of the data from 2006 to 2011. We then computed the conditional probability for the Ideal Point distance for each actor pair in 2012 given the distribution of action categories and determined the probability of various intervals around the true value. We considered different size intervals (5%, 10%, and 20%), and observed that our learned model computes conditional probabilities (6.9%, 13.7%, and 26.2%) that exceed the baseline predictions from a uniform distribution. The prediction here is obviously very noisy, but again, it is a very encouraging sign that a purely data-driven modeling algorithm can identify an informative dependency between only a small subset of event types (e.g., those in the Markov blanket) and UN voting patterns.

6 Analysis of Probabilistic Dependencies

Even if there is no direct link between two nodes in the Bayesian network, there can still be an indirect probabilistic dependency. In this section, we analyze networks over just GDELT events, without Ideal Point distances, allowing us to use an additional two years of data for which Ideal Point data is not available. Thus, we still use the same 133 event types, but now over the data of 9 years, 2006–2014 from GDELT.

For each pair of events, A and B, we query the learned Bayesian network to compute two conditional probabilities, $\Pr(B > \text{median}(B)|A > \text{median}(A))$ and $\Pr(B < \text{median}(B)|A < \text{median}(A))$. The former (latter) represents the likelihood that events of type B occur with high (low) frequency when events of type A occur with high (low) frequency. We can thus roughly characterize the *impact* of A on B by the difference between these two conditional probabilities. In other words, the greater the difference in the probability, the greater impact the occurrence of A events has on the likelihood of B events. We examined the learned Bayesian networks over the 9 years and identified 626 event pairs (out of 8778 possible) that had the same direction of impact across all of them. Table 6 shows the event pairs in that set with the highest impact.

Mean	**First Event**	**Second Event**
0.8425	Host a visit	Make a visit
0.1943	Allow international involvement	Provide military aid
0.1379	Allow international involvement	Provide military protection
0.1365	Mobilize armed forces	Provide military aid
0.1323	Mobilize armed forces	Allow international involvement
0.1291	Provide military aid	Express intent to accept mediation

Table 6. Highest impact event pairs For GDELT.

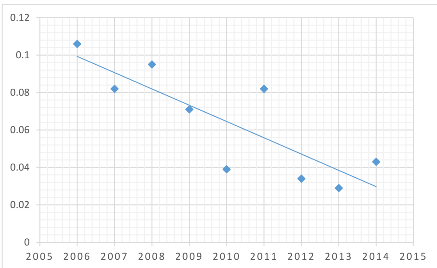
7 Identifying Anomalous Events

Table 6 shows that the impact for "Make a visit $\leftrightarrow$ Host a visit" is more than quadruple the next highest value. While this result is rather intuitive (i.e., when I make a visit to you, you host a visit for me), we wished to confirm the accuracy of our intuition. To do so, we manually reviewed 58 of the news links contained in the GDELT event records to informally verify the event. As it turned out, 51 of the 58 links that were categorized as "Make a visit" were also categorized as "Host a visit". Again, this would seem as expected, but further reading of the text revealed that only 14 of the links were actually categorized correctly, while the rest were not related to either making or hosting a visit.

Similarly, the 30 strongest negative impacts all included "Mass expulsion" as one of the events, despite the relative infrequency of mass expulsions over the last decade. Examining 51 of the source news articles categorized as "Mass Expulsion", we noticed that only 13 were relevant. While this partial investigation is not necessarily conclusive, it suggests an error in the parsing of these particular event types. Fortunately, manual inspection of the news source articles showed that such systematic errors are the exception in GDELT's extraction. However, the two examples found here demonstrate an ability of our methodology to unearth such anomalies in GDELT's extraction process.

8 Dynamics of Event Interdependency

Having already examined the consistency of the learned networks and their structure, we can also examine the consistency of event interdependency within our networks by analyzing changes in the impact that event types have on each other. By treating each Bayesian network as a summarization of the data from its given year, we can extract a time series of dependency impact values. A linear regression of the impacts over time for pairs of event types reveals interesting trends in terms of how the impacts are changing over time. For example, Figure 1a shows the change (or lack thereof) in the dependency between “Sign formal agreement” and “Express intent to cooperate”. Thus, not only is the occurrence of these two event types interdependent across actor relationships, but the magnitude of that interdependency has shown to be stable over the 9 years of data. In contrast, Figure 1b shows that the dependency between “Diminish military engagement” and “Provide military aid” has been weakening over time.



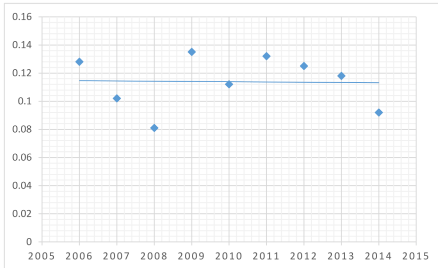


Fig. 1. (a) Impact between “Sign formal agreement” and “Express intent to cooperate”. (b) Impact between “Diminish military engagement” and “Provide military aid”.

9 Conclusion

In this work, we apply Bayesian network learning algorithms to available data on behavior at the geopolitical level. The link structure generated between different categories of actions provided a clear benefit in the explanation power of the models. Furthermore, the learned structures reveal qualitative properties of the relationships among action categories that can spur further investigation by political scientists.

It is important to note that, while we have limited our exploration to pairwise relationships in this paper, the Bayesian network’s representation of the underlying joint distribution allows us to measure the impact of arbitrary subsets of events on other arbitrary subsets. By leveraging this representation, we can thus greatly expand the space of possible queries that can be answered. The generality of this underlying AI model and its algorithms should empower political scientists to conduct analyses that are difficult through purely statistical methods.

The analysis of our models’ actor-specific performance showed that more data led to better accuracy. Exploiting additional data sources improves our models’ accuracy. Use of Bayesian network learning and inference algorithms makes it straightforward to incorporate such additional data. In fact, we applied our method to the iData repository of Integrated Crisis Early Warning System (ICEWS), which also uses CAMEO codes. The interesting observation was that the

Mean	First Event	Second Event
0.2530	Conduct strike or boycott	Protest violently, riot
0.2444	Conduct strike or boycott	Coerce
0.1634	Physically assault	Employ aerial weapons
0.1625	Conduct strike or boycott	Meet at a third location

Table 7. Highest impact event pairs for ICEWS.

list of the highest impact pairs were totally different for ICEWS data, even though the methods used were the same as the ones for GDELT data, see Table 7. This observation means that we could benefit from aggregation of two data sets, in order to have a more robust prediction of the highest impact pairs. Because each data set has its own types of inaccuracies, the two data sets could potentially complement each other's shortcomings. Thus, complementary sources like ICEWS promise to increase the accuracy of our models without any change in methodology.

There are potential limits to how accurate our purely data-driven models can be. It can be impossible to distinguish some international relationships based on just event counts extracted from the news. For example, Iran's relationship with both Argentina and Israel exhibit similar percentages of "Disapprove", "Accuse", and "Reject" events, yet the two relationships would be considered very different from a political point of view. It is likely that we may need to introduce domain knowledge from the political science literature. Such domain knowledge may come in the form of hidden variables or prior structures for our Bayesian networks from which our algorithms can bootstrap.

While there remains much more work to be done, our methodology here represents an important first step toward automatically learning computational models of international relations. The ever-increasing volume of online data offers a detailed source of geopolitical behavior that can move formal modeling beyond the high-level abstractions that have been necessary in the past. With the accompanying advances in AI algorithms for constructing such models from data, there is now a valuable opportunity for a new dialog between AI researchers and political scientists.

Acknowledgments

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). The authors also thank Torsten Woertwein, and Drs. Marco Scutari, Kalev Leetaru, Eric Voeten, Arthur Spirling, Jeffery Freiden, and Steven Brams.

References

1. Cimbala, S. J.: AI and National Security. Lexington Books. (1986)
2. Heckerman, D.: A tutorial on learning with Bayesian networks. Springer. (1998)
3. Hudson, V. M., ed.: AI and international politics. Westview Press. (1991)
4. Jensen, F. V.: An introduction to Bayesian networks, volume 210. London, UK, UCL Press. (1996)
5. Neapolitan, R. E.: Learning Bayesian networks. Prentice Hall. (2004)
6. Pearl, J.: Probabilistic reasoning in intelligent systems. Morgan Kaufmann. (1988)
7. Pearl, J.: Causality; models, reasoning and inference, volume 29. Cambridge University Press. (2000)
8. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann. (2014)
9. Schrodt, P. A.: AI and formal models of international behavior. The American Sociologist 19(1):7185. (1988)
10. Snidal, D.: The game theory of international politics. World Politics 38(01):2557. (1985)
11. Sylvan, D., and Chan, S.: Foreign Policy Decision Making; Perception, Cognition, and AI. New York, NY; Praeger. (1984)