

Use of force and civil–military relations in Russia: an automated content analysis

Brandon M. Stewart and Yuri M. Zhukov*

Department of Government, Harvard University, Cambridge, MA, USA

Russia's intervention in the Georgian–South Ossetian conflict has highlighted the need to rigorously examine trends in the public debate over the use of force in Russia. Approaching this debate through the prism of civil–military relations, we take advantage of recent methodological advances in automated content analysis and generate a new dataset of 8000 public statements made by Russia's political and military leaders during the Putin period. The data show little evidence that military elites exert a restraining influence on Russian foreign and defence policy. Although more hesitant than their political counterparts to embrace an interventionist foreign policy agenda, Russian military elites are considerably more activist in considering the use of force as an instrument of foreign policy.

Keywords: content analysis; document classification; civil–military; use of force; Russia; elites

On 14 August 2008, two days after ordering a halt to military operations in Georgia, Russian President Dmitry Medvedev addressed a group of senior military officers who had just returned from the conflict zone. 'We will call things by their names: the people of South Ossetia have survived genocide – and we should talk about it in this way. It will take years, maybe decades, to heal these wounds. And the fact that the annihilation of an entire nation was averted is lawful, unavoidable and absolutely justified, not to mention that the defense of its people, its citizens, is the direct duty of Russia as a state.'¹ Part humanitarian intervention, part self-defense, Medvedev's rationalization of Russia's first major use of force outside its borders since the Soviet–Afghan War has left many outside observers wondering about the future role of the military as an instrument of Russian foreign and defense policy. Whether one interprets the events of August 2008 as a manifestation of interventionism or realpolitik, the Russian president has characterized the use of force in either case as legitimate and necessary. What is less clear, however, is whether such a seemingly broad interpretation of the military's mission is widely shared by members of Russia's national security establishment.

*Corresponding author. Email: zhukov@fas.harvard.edu

In this study, we examine the debate over the use of force in Russia through the prism of civil–military relations. Do Russian political and military elites share the same foreign and defense policy priorities? Do they share the same attitudes toward the use of force? What explains convergence and divergence in these views? In addressing these three questions, we hope to lend new insights into recent patterns of Russian strategic thought and identify potential areas of consensus and conflict in Russian foreign and defense policymaking.

An emerging conventional wisdom among scholars of civil–military relations in the West has been that military elites tend to exert a restraining influence on foreign and defense policy. Compared to civilian policymakers, military professionals are expected to endorse a less expansive set of foreign policy priorities and a narrower set of missions for the armed forces. While the military conservatism hypothesis has inspired a wealth of academic debate in the US, few efforts have been made to extend this discourse to non-Western political systems, or to study intergroup opinion cleavages directly.²

We take stock of attitudes toward the use of force in Russia by (1) extending the ongoing conservatism–militarism debate to another type of political system, and (2) taking advantage of recent methodological advances in politically oriented automated content analysis. In particular, we employ supervised and unsupervised machine learning techniques to discern trends in the public debate over Russian foreign and defense policy, and offer a summary of data generated from 7920 public statements by political and military elites between 1998 and 2008.³ In so doing, we provide a data-driven characterization of contemporary civil–military relations.

Our analysis of these public statements shows little evidence of military conservatism in contemporary Russia. Although military elites are significantly less likely than political elites to embrace an interventionist foreign policy agenda, they also display attitudes that are considerably more permissive toward the use of force as an instrument of foreign policy. We also find that military and political foreign policy views tend to converge when the two groups of elites are more integrated in organs of executive power.

We begin by briefly reviewing the literature on civil–military relations and the use of force, and derive a series of descriptive hypotheses. Second, we introduce the methods employed to collect and analyze Russian text data. Third, we discuss our new dataset and the implications and limitations of our approach. We conclude with a summary of our findings and offer directions for future research.

To a hammer, everything looks like a nail?

Two views have emerged in the literature on political and military attitudes toward the use of force: military conservatism and military activism. The first of these holds that – compared to civilian policymakers – military elites are inclined to endorse a less expansive set of foreign policy priorities and a narrower

set of missions for the armed forces.⁴ This divergence in views stems in part from military organizational interests, which dictate that an expanding mission set, lengthy operational deployments and extended force postures undermine the readiness and capability of the armed forces to perform their core duties – to fight and win the nation's wars and safeguard its critical national interests. As a result, military officials are expected to be more reluctant than their civilian counterparts in recommending military action where core national interests are not perceived to be at stake.⁵ The use of force for humanitarian relief, regime change, crisis response and other interventionist purposes are areas where political and military opinions are expected to diverge.⁶

The second school of thought – military activism or militarism – sees military elites as more war-prone than their civilian counterparts. The military profession is seen as attracting a certain type of individual, who is subsequently socialized to prefer rapid decision-making over careful deliberation, and place excessive faith in the use of force as a solution to political problems – to a hammer, as the saying goes, everything looks like a nail.⁷ Even if perceptual biases are discounted, military organizational factors such as offensive doctrines and rigid mobilization timelines may raise the likelihood of conflict escalation from the diplomatic sphere to the military.⁸

While the academic discussion in the US has been rich and energetic, for the purpose of our study, it has fallen short in two areas. First, the empirical literature that it has produced has been largely confined to the United States and other Western democracies.⁹ Hypotheses derived from military conservatism literature have generally not been subject to rigorous empirical tests, while the conventional wisdom among scholars of Russian military affairs has tended to support the military activism position.¹⁰

Second, although many of the claims made on both sides of the debate rest on the existence of opinion gaps between political and military elites, few quantitative studies have sought to measure these gaps directly. Instead, authors have generally assumed the existence of such a gap and estimated its effect on decision-making with proxy variables, such as the military experience of policymakers or subjective measures of civilian control over the armed forces.¹¹

We address this empirical problem by estimating opinion cleavages through the analysis of political and military texts. Though largely absent from the literature on US civil–military relations, content analysis has had a long tradition in the study of Russian and Soviet foreign policy.¹² While attitudes over the use of force have not been at the centre of this literature, most Soviet-era studies found broad convergence in political and military views on questions of grand strategy, and minor levels of disagreement on lower-order issues like defense spending and procurement, where military institutional interests were more directly at stake.¹³ Soviet military elites were found to be less outspoken on foreign policy issues overall, although in some areas – such as support for national liberation movements in the Third World – they were sometimes found to be even more supportive of an interventionist role for the Soviet Union than

their civilian counterparts.¹⁴ The extent to which these Soviet-era patterns have persisted remains an open empirical question, although surveys conducted in the 1990s have found a clear and broad gap between civilian and military elite opinion on the source and magnitude of external threats.¹⁵

Hypotheses

As stated at the outset of this study, we are interested in answering three questions: Do political and military elites in Russia share the same foreign and defense policy priorities? Do they share the same attitudes toward the use of force? What explains convergence and divergence in these views? The following section outlines why we consider these questions relevant, and what we might expect to find in each case.

Issue salience

Issue salience can be understood as the relative importance a political actor assigns to a given policy area. The level of attention devoted to an issue is measured in part by the frequency with which it is raised and, insofar as this frequency distribution is not uniform across a population, issue salience can be said to reflect differences in policy priorities and domains.¹⁶ If, for instance, Mr Blue has a tendency to mention ethnic Russians in Crimea in every speech and never mentions missile defense systems, while General Red speaks only of missile defense systems and never of the Russian Diaspora, one can assume that Mr Blue sees the Diaspora as a more salient issue than missile defense, and General Red sees the opposite.¹⁷

Following the framework provided by Gelpi and Feaver (2004), we divide issues relevant to foreign and defense policy into two categories: *realpolitik* and *intervention*. The first category is more restrictive and reflects only those issues that represent a potentially substantial threat to state sovereignty. These include, but are not limited to, territorial defense, border security, nuclear and conventional deterrence, domestic separatism, protection of sea lines of communication, defense of allies and geostrategic access. The second category is more expansive and includes issues which do not necessarily impinge upon core national security interests as defined by realists, but 'may require intervention inside the boundaries of an allegedly sovereign state and may challenge the claims of sovereignty made by the ruling group within that state.'¹⁸ These include, but are not limited to, humanitarian assistance, foreign internal defense, peace operations, state building, crisis response, regime change and support for insurgencies in foreign states.

If the military conservatism hypothesis holds, the range of issues addressed in public statements by contemporary Russian military elites should be expected to reflect the *realpolitik* category. Statements by political leaders, meanwhile, can be expected to reach beyond these traditional national security issues.

- *H1: Military elites are less likely than political elites to find interventionist issues salient.*

Use of force

Although it can lend insights into the policy priorities of political and military elites, issue salience alone cannot fully capture actors' preferences for the employment of various instruments of state power. As empirically observable phenomena, attitudes toward the use of force are far more elusive than a simple measure of how often a given issue is raised. Government officials typically show great restraint in making statements about the use of coercion, choosing words carefully and permitting multiple interpretations without committing the government to a given action.¹⁹ Judgments of the relative hawkishness of political and military elites are thus inevitably subjective readings of whether the tone and content of a given statement could be considered consistent with an inclination toward the employment of military force.²⁰

While we admit that any dichotomous classification of attitudinal differences risks oversimplification, for the purpose of our second research question we assume perceptions of the utility of force to oscillate between two ideal types – conservatism and activism. Statements consistent with a conservative outlook display a preference toward the use of non-military instruments of power, unless all other means have been exhausted and a successful policy outcome is a near certainty.²¹ Such statements might include explicit doubts about the efficacy of military solutions to specific political problems, or more nuanced expressions of preferences – such as emphases on defensive, rather than offensive military capabilities, inclinations toward multilateral solutions for potential interventionist crises, and general support for limited ends and means in foreign policy pursuits.

Statements consistent with an activist outlook advocate limited constraints on the employment of the military instrument of power. While military force need not be the instrument of first choice, the language used would reflect the author's expressed opinion that it is an equally legitimate and effective means to achieve a desired policy outcome. Such statements might explicitly advocate the use of military power in a specific scenario, but are more commonly limited to expressions of support for offensive capabilities and power projection, unilateral solutions to interventionist crises, and a general preference for foreign policy outcomes that revise, rather than reinforce the status quo. Given the idiosyncrasies of individual opinions, conservatism and activism are of course somewhat arbitrary categories to which exact conformity can only be forced. Yet for our limited purpose of tracking general tendencies in elite preferences, they are sufficient.

If the military conservatism hypothesis holds, public statements by contemporary Russian military elites should be expected to be consistent with conservative attitudes on the use of force. However, such an expectation would be at odds with the conventional wisdom among observers of Russian military

affairs, which has characterized senior Russian officers as reliable foreign policy hawks.²² In light of this theoretical contrast between American and Russian civil–military relations, our second hypothesis assumes activism to be the expected pattern in Russian military statements.

- *H2: Military elites are more likely than political elites to support the use of force.*

Differences in views

The dynamics of convergence and divergence in issue salience and support for the use of force can be seen as indicative of the cohesiveness of political–military relations. A key determinant of group cohesion identified by military conservatism scholars has been the degree of elite integration, particularly the scope of formal authority that active or retired military elites might enjoy in political decision-making structures.²³ When a large proportion of political leadership has a professional military background, the two elites can be considered deeply integrated, and disagreements on foreign and defense policy are expected to decrease. If the military conservatism hypothesis holds, convergence and divergence in the views of Russian political and military elites would be associated with group integration.

- *H3: Military and political views converge when elites are closely integrated.*

Research design

To test these hypotheses, we perform document classification on an original dataset of 7920 speeches, press briefings, articles, interviews and other public statements made by members of Russia's political or military elite between 12 February 1998 and 31 October 2008.²⁴ This range of dates is intended to encompass both terms of the Putin presidency, along with two 'buffer' time periods on each end. Military elites are defined here as active duty and retired officers of field grade and above, or who – at the time of the statement's publication – occupied a position of formal authority or informal influence in the Ministry of Defense or an affiliated agency. Similarly, political elites include civilians in a position of formal authority or informal influence in the executive or legislative branches of the government of Russia, or an affiliated civilian organization.²⁵ As defined here, authority is derived from actors' respective positions in a defined social structure, while influence is derived from control over other sanctions and rewards not associated with the occupancy of a particular position in a formal structure.²⁶ In our sample, formal authority is assumed for persons employed in government agencies and ministries, and informal influence is assumed for individuals at affiliated research and policy analysis organizations, such as the Council on Foreign and Security Policy and the Academy of Military Sciences.

Two different methods of automated content analysis are employed in this study: unsupervised topic modeling and supervised document classification. These correspond respectively to indicators for issue salience and support for the use of force.

Bag-of-words analysis

Bag-of-words (BOW) is the fundamental assumption underlying the unsupervised topic models and supervised document classification methods. The assumption, common in the computational linguistics community, treats documents as a function of the words they contain, independent of the ordering or structure of the document.²⁷ While the bag-of-words assumption is admittedly unrealistic, the result is an efficacious model that has been shown to reasonably handle a variety of classification tasks.

Each analysis begins with a sparse document-term matrix ($d \times t$), which represents the set of documents to be analyzed in terms of their vocabulary. In order to achieve some dimensionality reduction, almost all BOW techniques make use of a stemmer or a lemmatizer to reduce words to their base meanings. Stemmers remove the ends of words in English to bring similar words down to one base meaning. For example, the words 'great, greater, greatest and greatness' are all reduced to the stem 'great-', and the words 'economy, economic, economical' are all reduced to the stem 'econom-'.²⁸

After the stemming process, a variety of statistical techniques can be applied to achieve further dimensionality reduction on the sparse matrix with the reduced ($d \times t^*$) matrix serving as the feature set. Thus, the only part of the process that is not language agnostic is the initial step of stemming and text processing.

We made the decision to analyze the documents in their native Russian.²⁹ Machine translation is notoriously unreliable, despite the relatively relaxed assumptions of BOW analysis. In order to provide the best results, words must be consistent in their meaning across texts. This requires that multiple words meaning different things in Russian not be translated to the same word in English. Unfortunately, machine translation is not even up to simple word-for-word translation that ignores lexical ordering.

From this common foundation, our approaches for coding the documents diverge. Supervised and unsupervised learning represent two different approaches in the field of machine learning. Supervised learning uses a training set of predefined categories to infer a technique for classifying texts according to the specified structure. Unsupervised learning uses no training set and seeks to cluster on the underlying structure of the documents.

Unsupervised topic models

Unsupervised learning methods are techniques for analyzing large sums of data using relatively few assumptions.³⁰ In this particular case we want to explore the range of topics to which Russian actors in the political and military bureaucracy

choose to devote their attention, with minimal a priori assumptions about the range and nature of those topics. For this task, we choose to adopt Grimmer's (2009) Expressed Agenda Model.

The model assumes a collection of documents that contain a natural hierarchical structure, with actors at the top of that hierarchy and the topics they discuss on the bottom. This structure is typically ignored in unsupervised learning problems. The model – described as an *Expressed Agenda Model* because it captures the attention that actors choose to publicly commit to a topic, rather than attempting to estimate private attention through actions – is designed for situations where the ‘quantities of interest are the priorities a set of actors allocate to issues’.³¹

The model requires the assumption of some number of k topics. The number of topics (k) is selected by the specification that yields the most interesting and substantively relevant topics, following common practice in the unsupervised learning literature.³² The topics are then labeled as either: (1) ‘Realpolitik’ – a category driven by words such as ‘counterweight’ (*protivoves*), ‘surpass’ (*operezhat*) and ‘military readiness’ (*boegotovnost*), (2) ‘Interventionist’ – driven by words such as ‘police’ (*politseyskiy*), ‘ethnic’ (*etnicheskiye*) and ‘freedom’ (*svoboda*), or (3) ‘None of the Above’.³³

Supervised document classification: an ensemble approach

Whereas unsupervised topic models are sufficiently sensitive for the clustering of documents by issue, our use of force variable necessitates classification of documents by a more specific type of expressed opinion (‘Activist’/‘Conservative’). To this end we employ a supervised learning approach, which uses a training set of 300 randomly selected, hand-coded statements to learn to classify the remaining documents.

In the training set, documents are coded ‘Activist’ if they express support for (1) the use of military power in a specific scenario, (2) the use of military power as the most effective means to reach concrete political ends, (3) unilateral solutions to international security crises, or (4) foreign policy outcomes that revise, rather than reinforce the status quo.³⁴ Documents consistent with a ‘Conservative’ classification include expressions of a preference for (1) the use of military power only as a last resort, (2) interagency and/or (3) multilateral solutions to international security crises, (4) limited ends and means in foreign policy, or (5) explicit doubts about the utility of military power in a specific scenario. All other documents are coded ‘None of the Above’. Following this typology, an example of an activist statement might be ‘Russia historically has been and will remain the guarantor of security for the peoples of the Caucasus [*Rossiya istoricheski byla i ostanetsya garantom bezopasnosti narodov Kavkaza*]’, whereas a conservative statement may be ‘One should resort to military power when the capabilities of other instruments are exhausted [*K voyennoy moshchi sleduyet pribegat*], *kogda vozmozhnosti drugih sredstv ischerpany*].’

In selecting our approach, we focus on improving what we view as a key deficiency of the literature as imported into political science – the failure to take into account the inherent inaccuracies of the classification process.³⁵ We seek to address

this problem using an ensemble learning approach, the combination of a variety of machine learning systems that function together to produce a classification.

Ensemble classifiers are optimal in situations where there is significant diversity in the individual classifiers' classification techniques and resulting feature space.³⁶ In these situations, ensemble classification – while not guaranteed to produce better classification rates than any individual classifier in the ensemble – provides increased stability on additional out-of-sample tests. Thus while out-of-sample testing using cross-validation can provide an assessment of system accuracy, the use of ensemble classification gives us confidence that those assessments are approximately accurate across the entire document set. Ensemble classification also provides us with a better picture of the uncertainty that exists within any single classification decision. The weighted predictions of each classifier can be interpreted as probabilities of each category for a given document.

Individual classifiers

We employ four different machine learning approaches in our ensemble classifier, two of which are ensemble classifiers in their own right: K-Nearest Neighbor, Adaboost.M1 algorithm, Random Forest and Support Vector Machine. Figure 1 shows how these classifiers are combined within the ensemble classification system.³⁷

- (1) *K-Nearest Neighbor* proceeds from an intuitive assumption that an unknown case can be classified according to its k nearest neighbors in the parameter space (in this case using Minkowski distance). We selected $k = 5$ and chose an un-weighted version of the model, where all five neighbors were treated equally.³⁸
- (2) *Adaboost.M1* uses several instances of a WeakLearner (in this implementation a classification tree) to generate hypotheses using data randomly drawn from the training distribution. The distribution is then iteratively updated to include instances misclassified by the first algorithm. This proceeds until a weighted majority vote occurs, which yields the final classification.
- (3) *Random Forest* is similar to the Adaboost algorithm but uses a different approach. It is also an ensemble technique and uses classification trees as its sub-component. Rather than iteratively training near examples missed by the classifier previously, the trees are grown using bootstrapped versions of the data and by choosing k nodes for which to search for a split. This introduces random perturbations into the data which generate different results in each tree and prevents over-fitting.³⁹
- (4) *Support Vector Machine (SVM)* is easily the most popular machine learning algorithm in political science due to its easy implementation and broad utility compared to other techniques.⁴⁰ SVM fits a hyper-plane to the feature space, which separates two categories of points from each other and maximizes the marginal distance between the nearest points and the surface.

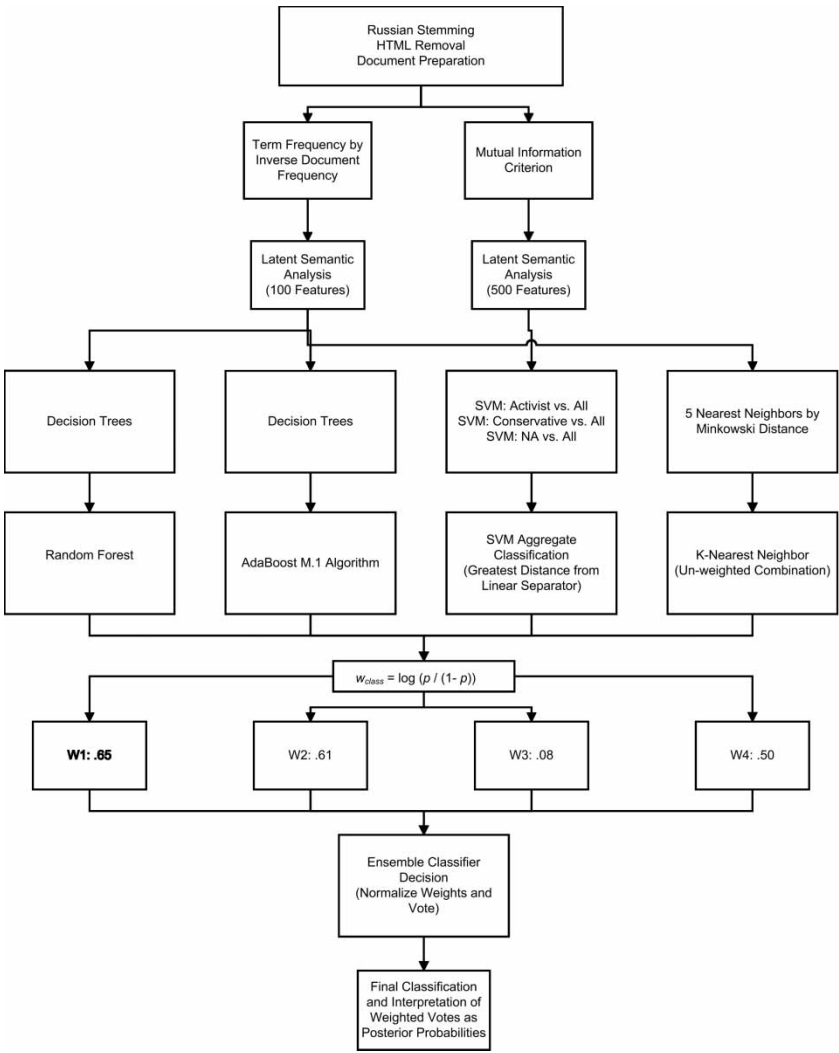


Figure 1. Diagram of ensemble classification system.

The ensemble

In order to effectively weight the predictions from the classifiers, we obtain accuracy measures by cross-validation for each algorithm. Using our training set of 300 hand-coded documents, we randomly sample 275, train the algorithm and then test on the out-of-sample 25 documents. We repeat this simulation 10,000 times for each algorithm to attain out-of-sample accuracy results. We then weight the classifiers predictions based on those results.

Given these weights, we simulate the training of each classifier on the same sample of 275, apply weights to the predictions, attain a final prediction and then verify against the out-of-sample 25. We repeat this process 10,000 times to produce the accuracy rates for the ensemble system as a whole. The accuracy of the system is lower than any individual classifier, but not dramatically. Nonetheless, the ensemble approach guards against the idiosyncrasies of any given classifier.

Uncertainty

We use the estimated weights to incorporate the uncertainty of the classification procedure into our analysis, a form of uncertainty which is often discarded. We infer probabilities of categorization from the individual weights, and then redraw the data 10,000 times using a distribution defined by the probabilities for each class and observation. We can then use this sampling distribution of documents to calculate our quantities of interest, without introducing bias from the classification error.

While all variables have some degree of measurement error, the advantage of our application of an ensemble classifier is the ability to specifically quantify that error. To create the estimates of document classification by category, we take the mean value within each category of interest across all 10,000 runs. All confidence intervals throughout the paper are calculated by redrawing from our classification probabilities and rerunning the analysis to provide a full range of estimates possible under our classifier.

Summary

A summary of document classification is shown in Figure 2. The left pane shows the output of the unsupervised topic model (issue salience), while the output of the ensemble classifier (use of force) is shown in the right pane. Over half of the documents in the dataset are classified ‘none of the above’ for issue salience, while the remainder is roughly evenly split between ‘interventionist’ and

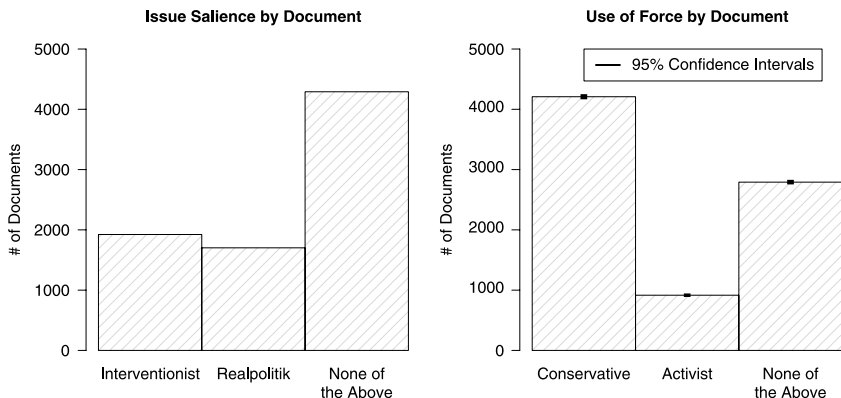


Figure 2. Summary of document classification.

‘realpolitik’ topics. Meanwhile, the number of documents which express a conservative position on the use of force is more than four times higher than those classified as activist. As indicated by the 95% confidence bands in the plot, the margin of error is quite small relative to the number of documents in each category, suggesting that our automated classification approach is sufficiently accurate to permit the drawing of inferences from these quantities.⁴¹

The Russian public debate: empirical findings

Our findings suggest that military conservatism has not always characterized the public debate over foreign and defense policy in contemporary Russia. While military elites are less likely than political elites to find interventionist policy issues salient, they are also far more likely to express an activist view on the use of force. This pattern appears to hold for persons in positions of both formal authority and informal influence. Further, political–military disagreement over foreign and defense policy appears to decline when the military brass is integrated in political decision-making structures.

Realpolitik and interventionism

Evidence of a civil–military opinion gap is quite apparent from Figure 3, which presents comparative distributions of average monthly issue salience and activism scores between the two groups. The left pane of the figure suggests that statements by Russian political elites have been far less likely than those by military elites to address realpolitik foreign policy issues. Consistent with the military conservatism hypothesis, the distribution of political elite statements (solid line) has a mode near 1.0, representing interventionist issue salience, while the distribution of military elite statements (dashed line) has thicker tails and a mode around 0.15, much closer to the realpolitik end of the spectrum. Surprisingly, military elite statements appear to cover a relatively wide range of topics. While in any given month, interventionist

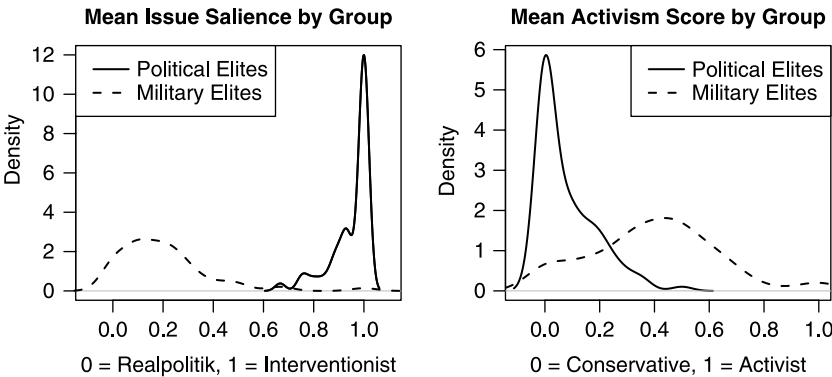


Figure 3. Distributions of dependent variables.

topics occupied the majority of statements by political elites – the left tail of the distribution does not dip below 0.6 on the scale – military elites did not avoid discussion of either type of foreign policy issue. Indeed, the distribution of military statements covers the entire range of the issue scale.

This finding is particularly striking in light of Soviet-era content analysis studies, which found a significant stay-in-your-lane pattern in public statements by military leaders. Insofar as a topic was considered ‘political’ or otherwise outside the scope of a military professional’s institutional portfolio, public statements on that issue by Soviet officers were exceedingly rare.⁴² To be fair, military support for interventionism during Soviet times was not unheard of – indeed, support for national liberation movements was among the articles of faith of the Marxist-Leninist foreign policy wholeheartedly embraced by the USSR’s national security establishment.⁴³ At the same time, it would be difficult to attribute the pattern shown on Figure 3 to residual elements of Soviet-era foreign policy dogma alone. Another explanation may be hierarchical cue-taking, wherein the rhetoric of military officers is to some degree influenced by informal cues or formal strategic guidance from political authorities. In expressing opinions that fall outside the scope of core military institutional priorities, military leaders may thus be signaling concurrence with the stated views of public officials or simply adhering to official state policy.⁴⁴

Curiously, contemporary Russian military elites appear far more eager to discuss non-traditional security issues like ethnic conflict, policing and freedom than their civilian counterparts are to discuss ‘traditional’ security issues like territorial defense, military readiness and the balance of power. Evidently, contemporary Russian political leaders seem no less deferent to military expertise on national security issues than their predecessors had been during the Soviet period.⁴⁵ The public and academic discourse on military security issues – as defined here by the *realpolitik* category – continues to operate as a mostly autonomous sphere, largely insulated from civilian perspectives.

A more detailed representation of issue salience is provided in Figure 4. This bar plot shows the proportion of interventionist and *realpolitik* statements by representatives of six groups: (1) military and (2) political elites in positions of formal authority, (3) military and (4) political elites in positions of informal influence, (5) members of the Russian Parliament, including the State Duma and Federation Council and (6) members of the Presidential Administration and cabinet-level officials in the government of Russia.

Consistent with Hypothesis 1, the bar plot suggests that statements by both formal and informal military elites are highly likely – 86.9% and 84.2%, respectively – to discuss *realpolitik* topics. Meanwhile, political elites – particularly in the legislative branch and in positions of informal influence – are far more likely to address interventionist issues. The group most likely to discuss *realpolitik* topics is formal military elites, which includes senior officers currently on active duty – precisely the group of elites for whom core military institutional interests are most pertinent.⁴⁶ The proportion of *realpolitik* statements is a bit

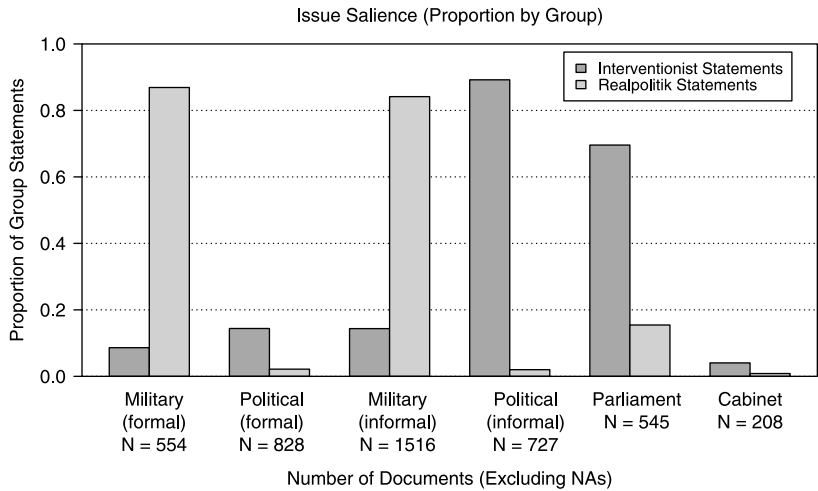


Figure 4. Interventionist and realpolitik statements by group.

lower among the retired senior officers included in the informal military category, who are presumably less constrained by protocol and institutional conventional wisdom, and are slightly more likely to address interventionist topics.

However, even informal military elites find themselves at a great distance from their political counterparts. The proportion of interventionist statements is highest (89.2%) among political elites in positions of informal influence – a group that includes academicians, think tank research fellows, prominent columnists and former civilian government officials. Of statements by members of parliament, 69.6% address interventionist statements, although this group was also more likely than other political elites to consider realpolitik issues salient – 15.4%, compared with 2.2% and below for all other non-military groups. Finally, a relatively small proportion of statements by Cabinet officials and other formal political elites are devoted to either realpolitik or interventionism, suggesting that the bulk of statements by these groups addresses neither topic. Even so, the general trend is the same: political elites are more likely to address interventionist issues than realpolitik issues, while statements by military elites follow the opposite pattern.

Conservatism and activism

Expressed opinions on the use of force, shown in the right pane of Figure 2, are less polarized than the results of the topic model, but still suggest a clear distinction between the policy positions of civilian and military elites. Contrary to the military conservatism hypothesis, the distributions point to a relatively high level of permissiveness toward the use of force in statements by military elites. The distribution of political elite statements (solid line) has a mode near zero, indicating a generally conservative position on the use of military force.

The sentiments of military elites (dashed line) are clustered near the middle of the continuum – hardly monolithic in their support for the use of force, but nevertheless far closer to an activist outlook on average. As in the left pane, the military elite sample covers a relatively broad range of views, has a lower density in its mode and thicker, longer tails, extending to both extremes of the conservative–activist scale.

To the extent that military elites are more permissive toward the use of force than civilians, this figure seems to support the conventional wisdom among observers of post-Soviet military affairs, that military activism – coupled with political deference to military points of view – is the dominant pattern in Russian military opinion.⁴⁷ A number of scholars have argued that Russian military leadership tends to place excessive faith in the use of force as an instrument of conflict resolution, effectively limiting policy options available to national command authorities and occasionally facilitating the rapid escalation of conflicts. For instance, qualitative case studies of uses of force during the interventions in Transdnistr and Tajikistan have attributed decisions to employ military power to the initiative of local commanders, rather than to policy guidance from the Kremlin – which was either limited or non-existent at the time.⁴⁸ Although caution must be exercised in reading too deeply into the distributions shown here, the tone and content of military and political statements under study appear to reflect such a conceptualization of Russian military views.

Consistent with these trends, the bar plot shown in Figure 5 supports our second hypothesis.⁴⁹ While statements by military elites are on balance more likely to make a conservative rather than activist statement on the use of force, the proportion of activist statements is considerably higher among military elites than among their political counterparts. 24.4% of formal military and 26.4% of informal military statements convey attitudes consistent with an activist view toward the use force, compared with less than 10% for most

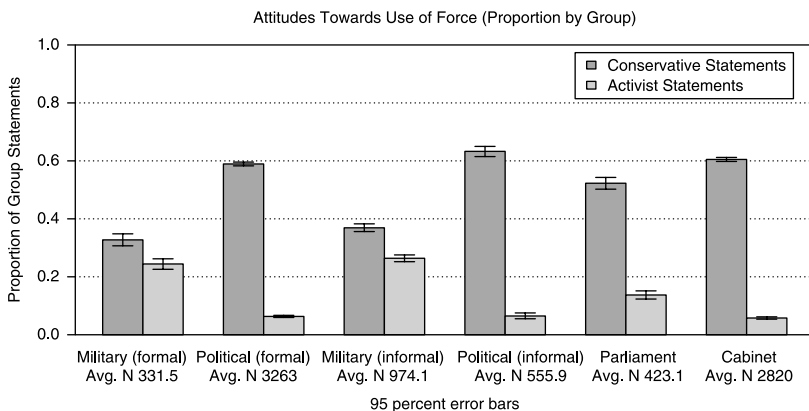


Figure 5. Conservative and activist statements by group.

political groups. This relative permissiveness in considering the use of force as an instrument of foreign policy applies to both formal and informal military elites, although the informal subgroup is slightly – but not much – more likely to express a conservative position. Statements by civilians, meanwhile, tend toward the conservative side of the spectrum. Among political elites, members of the State Duma and Federation Council have the highest proportion of activist statements, at 13.7%, while this statistic is less than 7% for all other political groups.

Agreement and disagreement

While Figures 3–5 offer a glimpse of how political and military views differ on issue salience and support for the use of force, the spatial scatterplot in Figure 6 displays differences in views on these two dimensions simultaneously, with monthly averages of group issue salience scores on the y-axis and use of force scores on the x-axis. A point in the lower left-hand corner thus indicates that a majority of statements in a given month are realpolitik and conservative, while a point in the upper right-hand corner indicates that the majority are interventionist and activist. As one might expect from the above discussion, political elite scores

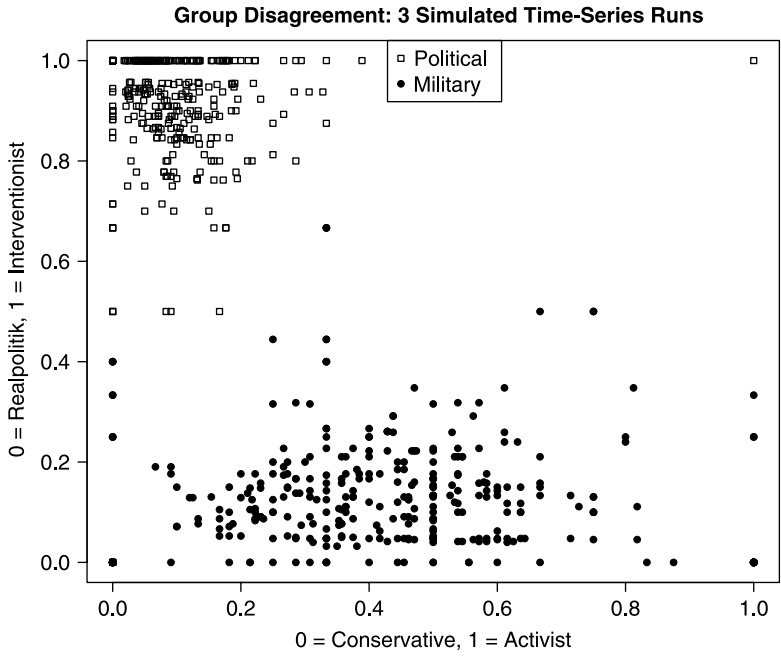


Figure 6. Spatial representation of average monthly issue salience and use of force score by group.

are clustered in the upper left-hand corner, suggesting that – most months – statements by Russian political elites addressed interventionist topics, but were relatively conservative with respect to the use of force. Military elite scores, meanwhile, are clustered near the bottom of the plot, indicating a far higher salience for realpolitik topics, and near the middle on the *x*-axis, indicating a relatively more activist position on the use of force.

Figure 6 further suggests that military and elite positions are not static across time. While, on some months, average military scores appear in the lower-right quadrant of the plot, at other times, the military group is far closer to the political cluster in the opposite corner, indicating relative similarity between the positions of the two groups. To capture this convergence and divergence in views, we use the Euclidean distance between average monthly political and military positions on issue salience and the use of force, where higher distances indicate greater levels of disagreement and smaller distances indicate the opposite.⁵⁰ This spatial measure of disagreement is shown on Figure 7 as a function of elite integration – in this case the proportion of active duty or retired military elites on the Russian Security Council.⁵¹

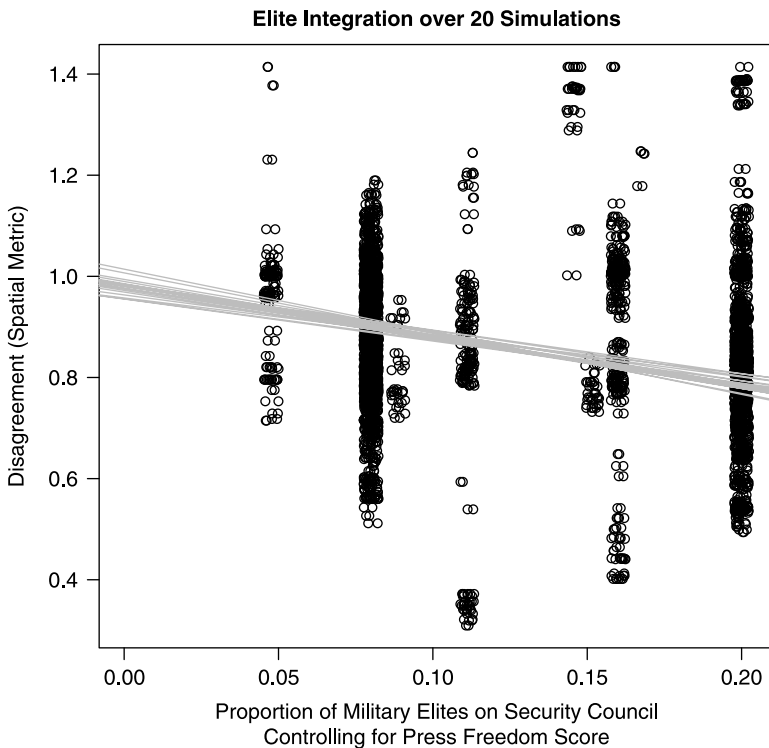


Figure 7. Elite disagreement and elite integrations.

Consistent with Hypothesis 3, Figure 7 suggests that political and elite views converge when the two groups are more tightly integrated in the highest echelons of the executive branch. As the proportion of Russian Security Council members with a professional military background increases, the level of disagreement falls. Indeed, some of the largest drops in the level of disagreement have historically coincided with increases in Security Council elite integration. One of these drops, from approximately 1.4 to 0.4, occurred after the election of Vladimir Putin as president and the subsequent appointment of prominent active duty and retired army generals – including Viktor Kazantsev and Konstantin Pulikovskiy – as presidential envoys to Federal Districts. Views converged again in 2001 and 2002, when the proportion of Security Council officials with a military background increased to 20%.

Despite high levels of integration during Putin's first term, disagreement spiked in the first half of 2003, when the president oversaw a partial restructuring of the national security bureaucracy, including the expansion of the powers of the State Security Service (FSB) and the establishment of a new anti-drug trafficking agency. Following these reforms, military elites on the Security Council were gradually displaced by security services professionals from the FSB, Foreign Intelligence Service and Ministry of the Interior. Elite disagreement rose slightly between 2004 and 2005, by which point the proportion of military elites on the Council had dropped by more than half. By this point in Putin's presidency, however, patterns of elite disagreement had largely stabilized, presumably reflecting a consolidation of power in the Kremlin, as well as tightening controls on mass media. More recently, levels of disagreement have followed a slightly downward path – a trend seemingly reinforced by the appointment of additional military elites to the Security Council after the election of Dmitry Medvedev as president in 2008.

Conclusions

Russian political and military elites differ on both the priorities of their country's foreign and defense policy and the role of military force as an instrument of national power. While military elites are less likely than political elites to attach salience to interventionist policy issues like ethnic conflict, policing and regime change, their views are also far more permissive toward the use of force as a policy instrument. These findings suggest that military conservatism – the view that military professionals are a voice of restraint in foreign policy – does not adequately reflect the dominant patterns of civil–military relations in the contemporary Russian state. We also find that the scale of the civil–military opinion gap in Russia depends on the extent to which the two elites are closely integrated in organs of executive power. When a higher proportion of officials on the Russian Security Council has had a military background, opinion cleavages have been less likely to emerge in the public debate.

In addition to the substantive conclusions, we show how the analysis of traditionally opaque civil–military relations can benefit from the introduction

of automated content analysis techniques. We demonstrate two different approaches to document classification, which have different strengths and underlying assumptions. Unsupervised classification allows for issue salience to be modeled with few *a priori* assumptions about the structure of those documents or the issues they might engage. Supervised document classification follows in the tradition of Soviet-era manual content analysis, while allowing not only for classification on a larger number of documents, but also a more precise estimation of measurement error.

Of course, an individual's expressed views on the use force and attention to a given issue may be driven by a host of considerations other than group affiliation. Building on the descriptive statistics presented here, future research would need to examine the Russian civil–military opinion gap in a multivariate setting, with control variables drawn from a variety of literatures. For instance, studies of diversionary war have expected support for the use of force to increase with the emergence of a perceived external threat and during periods of heightened domestic unrest.⁵² With respect to elite disagreement, the in-group/out-group hypothesis has linked increased policy consensus and internal solidarity with the incidence of international crises and the emergence of foreign threats.⁵³ The influence of these and other exogenous factors must be rigorously evaluated in any subsequent analyses of the Russian public debate.

Russia's intervention in the Georgian–South Ossetian conflict in August 2008 has highlighted the need to develop a broad empirical base of evidence about the priorities and preferences of Russian political and military elites. Such evidence can assist scholars and policymakers alike, by identifying areas of consensus and conflict in Russian foreign and defense policy, identifying bastions of support and opposition in a given policy domain, informing strategic communications and guiding the content of military-to-military engagement. In taking advantage of methodological advances in content analysis to address a largely descriptive set of questions – on issue salience, support for the use of force and disagreement – this paper represents but one effort to improve our understanding in these areas. Our limited inquiry has hardly offered an occasion for optimism – Russia's military leaders appear as hawkish as they had been during Soviet days, while political elites have remained deferential and relatively reluctant to become involved in a meaningful debate over national security policy. Still, we hope that future research will further improve our grasp of the determinants Russian elite opinion and help manage our expectations about the future course of Russian foreign policy.

Acknowledgements

The authors would like to acknowledge assistance and helpful comments from Muhammet Bas, Matt Blackwell, Jeff Friedman, Justin Grimmer, Gary King, Kyle Marquardt, Rich Nielsen, Dan O'Huiginn and Beth Simmons. Any remaining errors are the authors' own. Financial support for this research project has been provided by the Davis Center for Russian and Eurasian Studies at Harvard University. Replication data and Appendix with additional estimation details are available on request from the authors.

Notes

1. Medvedev, 'Vstrecha s voennosluzhashimi Vooruzhennykh Sil Rossi'.
2. Notable contributions to this debate include Betts, *Soldiers, Statesmen and the Cold War*; Brodie, *War and Politics*; Ekirch, *The Civilian and the Military*; Huntington, *The Soldier and the State*; Gelpi and Feaver, 'Speak Softly and Carry a Big Stick?'; Gelpi and Feaver, *Choosing Your Battles*; Sechser, 'Are Soldiers Less War-Prone than Statesmen?'; Vagts, *A History of Militarism*.
3. See Quinn et al., 'How To Analyze Political Attention With Minimal Assumptions and Costs'; Shulman, 'Editor's Introduction', for two excellent surveys of current work in political science.
4. This view builds on Huntington's, *The Soldier and the State* classic thesis on professional military ethic, which holds that 'grand political designs and sweeping political goals are to be avoided, not because they are undesirable, but because they are impractical. The military security of the state must come first.' Huntington, *The Soldier and the State*, 68.
5. See Betts, *Soldiers, Statesmen and the Cold War*.
6. See Gelpi and Feaver, 'Speak Softly and Carry a Big Stick?' and *Choosing Your Battles*. For a discussion of traditional and non-traditional military missions, see the recent special issue of *Small Wars and Insurgencies* 19, no. 3 (September 2008).
7. See Vagts, *A History of Militarism*; Ekirch, *The Civilian and the Military*; Brodie, *War and Politics*, 495.
8. See Snyder, *The Ideology of the Offensive*; Van Evera, *Causes of War*; Sechser, 'Are Soldiers Less War-Prone than Statesmen?'.
9. Notable exceptions have included Scobell, 'Soldiers, Statesmen, Strategic Culture and China's 1950 Intervention in Korea' and 'Show of Force', which have found empirical support for the military conservatism hypothesis in, respectively, China's 1950 intervention in Korea and the Taiwan Strait Crisis of 1995–1996.
10. Azrael and Payin, 'US and Russian Policymaking'; Golts, 'Bremya Militarizma'; Gomart, *Russian Civil-Military Relations*.
11. See Gelpi and Feaver, 'Speak Softly and Carry a Big Stick?'; Sechser, 'Are Soldiers Less War-Prone than Statesmen?'. Another alternative is opinion surveys. Gelpi and Feaver, *Choosing Your Battles* analyzed survey data on civilian and military attitudes to test the hypothesis that US civilian elites are more supportive of intervention than military elites.
12. Manual thematic analysis of speeches and articles by party and military elites has been used to gauge trends in leadership commentary over a range of substantive areas, from education policy (Stewart, 'Soviet Interest Groups and the Policy Process') to defense expenditures (Zimmerman and Palmer, 'Words and Deeds in Soviet Foreign Policy'), deception (Axelord and Zimmerman, 'The Soviet Press on Soviet Foreign Policy') and articulation of dissatisfaction with federal policies (Breslauer, 'Is There a Generation Gap in the Soviet Political Establishment?'). While the statistical approach to content analysis is new for this literature, the underlying analytic assumptions follow from this previous tradition.
13. Frost, 'A Content Analysis of Recent Soviet Party-Military Relations'.
14. Stewart, Warhola and Blough, 'Issue Salience and Foreign Policy Role Specialization'; Frost, 'A Content Analysis of Recent Soviet Party-Military Relations'; Wallander, 'Third World Conflict in Soviet Military Thought'.
15. Zimmerman, *The Russian People and Foreign Policy* and 'Slavophiles and Westernizers Redux'.
16. To be 'salient', an actor's statements on a given issue need not take the form of coherent policy proposals. These assumptions find precedent in previous work primarily in American politics on senate press releases (Grimmer, 'A Bayesian

- Hierarchical Topic Model') and senate floor debates (Quinn et al., 'How to Analyze Political Attention'; Monroe et al., 'Fightin' Words').
17. Such differences in expressed priorities can, of course, be dictated by bureaucratic portfolios, making it difficult to separate the salience of an issue from the number of personnel assigned to it. Yet since we are interested primarily in *expressed* opinions, it is of secondary importance whether a statement is informed by bureaucratic parochialism, privately held views, or some combination of the two. In either case, the statement conveys the relative importance a group (Military/Political) publicly assigns to a given policy area.
18. Gelpi and Feaver, *Choosing Your Battles*, 22–3.
19. Axelrod and Zimmerman, 'The Soviet Press on Soviet Foreign Policy'.
20. We can classify these subtleties by reference to the document's vocabulary, which is a function of tone and content. This model of text analysis is well established both in sentiment analysis (Godbole et al., 'Large-Scale Sentiment Analysis for News and Blogs'; Hopkins and King, 'A Method of Automated Nonparametric Content Analysis for Social Science') and bag-of-words classification (Manning and Schütze, *Foundations of Statistical Natural Language Processing*).
21. Huntington, *The Soldier and the State*.
22. Golts, 'Bremya Militarizma' and "'Grazhdanskiy kontrol'" po-putinnski'; Gomart, *Russian Civil-Military Relations*; Miller and Trenin, *Vooruzhennye sily Rossii*; Golts and Putnam, 'State Militarism and Its Legacies'.
23. Gelpi and Feaver, 'Speak Softly and Carry a Big Stick?'. For a discussion of contested institutions and civil–military relations, see Dassel, 'Civilians, Soldiers and Strife'.
24. The corpus of documents includes 7920 statements by 533 individuals in positions of formal authority and informal influence, published in government and independent periodicals. While the sample is neither random nor universal, it is a reasonable representation of the public debate. A full source list can be found in Appendix A (available on request from authors).
25. Full lists of all considered elites are available in Appendix B1.
26. Huntington, *The Soldier and the State*, 86–90.
27. Manning and Schütze, *Foundations of Statistical Natural Language Processing*, 237.
28. Only unigrams (single words) are considered. Full text processing details can be found in Appendix C1.
29. While Russian is more grammatically rigid and morphologically more complex than English, it is roughly as sensitive to word order as English, making the BOW assumption roughly equivalent across the two languages.
30. Quinn et al., 'How to Analyze Political Attention'.
31. Grimmer, 'A Bayesian Hierarchical Topic Model for Political Texts', 4.
32. Blei et al., 'Latent Dirichlet Allocation'; Grimmer, 'A Bayesian Hierarchical Topic Model for Political Texts'; Quinn et al., 'How to Analyze Political Attention'.
33. Clusters are often described by the words that most influence the placement of a document in a particular category. Further lists of key words are available in Appendix C3.
34. Examples of each type of statement are provided in Appendix B.
35. Whether in automated or manual document analysis, error rates or inter-coder reliability rates are rarely taken into account at the level of estimation. Due to our relatively small corpus of documents we are unable to adopt techniques like those described in Hopkins and King, 'A Method of Automated Nonparametric Content Analysis for Social Science', which rely on having tens of thousands of documents at a minimum.
36. Polikar, 'Ensemble Based Systems in Decision Making'.

37. Specific settings, feature extraction techniques and package for estimation are available in Appendix C4.
38. Hechenbiechler and Schliep, 'Weighted k-Nearest-Neighbor Techniques and Ordinal Classification'.
39. Breiman, 'Random Forests'.
40. Hillard et al., 'Computer Assisted Topic Classification for Mixed Methods Social Science Research'; Yang and Liu, 'A Re-examination of Text Categorization Methods'.
41. The 95% confidence intervals are the 5th and 95th quantiles of the distribution.
42. Frost, 'A Content Analysis of Recent Soviet Party-Military Relations'.
43. Wallander, 'Third World Conflict in Soviet Military Thought'.
44. The authors are grateful to Kyle Marquardt for this insight.
45. Colton, *Commissars, Commanders, and Civilian Authority*.
46. See Bondaletov, "'Sotsial'no-protestnaya aktivnost" voennosluzhashchikh'.
47. Golts, 'Bremya Militarizma' and "'Grazhdanskiy kontrol" po-putinnski'; Gomart, *Russian Civil-Military Relations*; Miller and Trenin, *Vooruzhennye sily Rossii*; Golts and Putnam, 'State Militarism and Its Legacies'.
48. Dubnov, 'Tadjikistan'; Selivanova, 'Trans-Dniestria'.
49. The bar plot shows expected probabilities and 95% confidence intervals derived from simulation.
50. The use of Euclidean distance allows us to measure disagreement between our groups on two different variables simultaneously. Operationalizing disagreement by simple spatial models is common in the study of parties (Poole, *Spatial Models of Parliamentary Voting*) and here does not make any additional scaling assumptions other than that the two dimensions are equally important.
51. The plot shows 20 simulated slope coefficients for a linear regression of disagreement on elite integration, controlling for annual press freedom scores (Freedom House, 'Press Freedom Scores: 2009 Edition'). By providing simulations across 20 iterations of the model, we not only demonstrate the 95% confidence interval of the model (the upper and lower most lines), but also show that the relationship is robust across multiple draws of the Use of Force Classification. Regression lines substantively show the effect on disagreement of increasing integration while holding Press Freedom at its mean.
52. Domke, *War and the Changing Global System*; Gelpi, 'Democratic Diversions'; Levy, 'Declining Power and the Preventative Motivation for War' and 'The Diversionary Theory of War'; Ostrom and Job, 'The President and the Political Use of Force'; Russett, 'Economic Decline, Electoral Pressure, and the Initiation of Interstate Conflict'; Skocpol, *States and Social Revolution*; Wilkenfeld, 'Domestic and Foreign Conflict Behaviour of Nations' and *Conflict Behaviour and Linkage Politics*; Zinnes and Wilkenfeld, 'An Analysis of Foreign Conflict Behavior of Nations'.
53. Coser, *The Functions of Social Conflict*; Mueller, *War, Presidents and Public Opinion*; Polsby, *Congress and the Presidency*; Simmel, 'The Persistence of Social Groups'; Waltz, 'Electoral Punishment and Foreign Policy Crisis'.

Bibliography

- Axelrod, Robert, and William Zimmerman. 'The Soviet Press on Soviet Foreign Policy: A Usually Reliable Source'. *British Journal of Political Science* 11 (1981): 183–200.
- Azrael, Jeremy R., and Emil A. Payin, eds. *US and Russian Policymaking With Respect to the Use of Force*. Santa Monica, CA: RAND, 1997.
- Betts, Richard. *Soldiers, Statesmen and the Cold War*. Cambridge, MA: Harvard University Press, 1977.

- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 'Latent Dirichlet Allocation'. *Journal of Machine Learning Research* 3 (2003): 993–1022.
- Bondaletov, Valeriy. 'Sotsial'no-protestnaya aktivnost' voennosluzhashchikh [Social protest activity of military servicemen]. *Voennaya Sotsiologiya*, no. 11 (November 2006).
- Breiman, Leo. 'Random Forests'. *Machine Learning* 45, no. 1 (2001): 5–32.
- Breslauer, George W. 'Is There a Generation Gap in the Soviet Political Establishment?: Demand Articulation by RSFSR Provincial Party First Secretaries'. *Soviet Studies* 36, no. 1 (1984): 1–25.
- Brodie, Bernard. *War and Politics*. New York: Macmillan, 1973.
- Colton, Timothy. *Commissars, Commanders, and Civilian Authority: The Structure of Soviet Military Politics*. Cambridge, MA: Harvard University Press, 1979.
- Cortes, Esteban Alfaro, Matias Gamez Martinez, and Noelia Garcia Rubio. *adabag: Applies Adaboost.M1 and Bagging*. R package version 1.1.
- Coser, Lewis A. *The Functions of Social Conflict*. New York: Free Press, 1956.
- Dassel, Kurt. 'Civilians, Soldiers and Strife'. *International Security* 23, no. 1 (1998): 107–40.
- Domke, William. *War and the Changing Global System*. New Haven: Yale University Press, 1988.
- Dubnov, Arkady. 'Tadjikistan'. In *US and Russian Policymaking With Respect to the Use of Force*, ed. Jeremy R. Azrael and Emil A. Payin. Santa Monica, CA: RAND, 1996.
- Ekirch, Arthur A. *The Civilian and the Military: A History of the American Antimilitarist Tradition*. New York: Oxford University Press, 1956.
- Freedom House. 'Press Freedom Scores: 2009 Edition', Dataset. www.freedomhouse.org, 2009.
- Frost, Howard E. 'A Content Analysis of Recent Soviet Party-Military Relations'. *American Journal of Political Science* 3, no. 1 (1989): 91–135.
- Gelpi, Christopher. 'Democratic Diversions: Governmental Structure and the Externalization of Domestic Conflict'. *Journal of Conflict Resolution* 41 (1997): 255–82.
- Gelpi, Christopher, and Peter D. Feaver. 'Speak Softly and Carry a Big Stick? Veterans in the Political Elite and the American Use of Force'. *The American Political Science Review* 96, no. 4 (2002): 779–93.
- Gelpi, Christopher, and Peter D. Feaver. *Choosing Your Battles: American Civil-Military Relations and the Use of Force*. Princeton, NJ: Princeton University Press, 2004.
- Godbole, Namrata, Manjunath Srinivasiah, and Steven Skiena. 'Large-Scale Sentiment Analysis for News and Blogs'. *Proceedings of International Conference on Weblogs and Social Media* (2007).
- Golts, Aleksandr. 'Bremya Militarizma [The Burden of Militarism]'. *Otechestvennye Zapiski* 25 (2005).
- Golts, Aleksandr. 'Grazhdanskiy kontrol' po-putinski [Civilian Control Putin-Style]'. *Ezhednevniy Zhurnal*, 23 February 2007.
- Golts, Alexander M., and Tonya L. Putnam. 'State Militarism and Its Legacies: Why Military Reform Has Failed in Russia'. *International Security* 29, no. 2 (2004): 121–58.
- Gomart, Thomas. *Russian Civil-Military Relations: Putin's Legacy*. Washington, DC: Carnegie, 2008.
- Grimmer, Justin. 'A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases' (working paper, 2009).
- Hastie, Trevor. *svmpath: the SVM Path algorithm*. R package version 0.92, 2006. <http://www-stat.stanford.edu/~hastie/Papers/svmpath.pdf>.

- Hechenbichler, K., and K.P. Schliep. 'Weighted k-Nearest-Neighbor Techniques and Ordinal Classification'. Discussion Paper 399, SFB 386, Ludwig-Maximilians University Munich, 2004.
- Hillard, Dustin, Stephen Purpura, and John Wilkerson. 'Computer Assisted Topic Classification for Mixed Methods Social Science Research'. *Journal of Information Technology and Policy* 4, no. 4 (2008).
- Hopkins, Daniel, and Gary King. 'A Method of Automated Nonparametric Content Analysis for Social Science' (working paper, 2008).
- Huntington, Samuel. *The Soldier and the State: the Theory and Practice of Civil-Military Relations*. Cambridge, MA: Belknap Press, 1957.
- Levy, Jack. 'Declining Power and the Preventive Motivation for War'. *World Politics* 40, no. 1 (1987): 82–107.
- Levy, Jack. 'The Diversionary Theory of War: A Critique'. In *Handbook of War Studies*, ed. Manus I. Midlarsky. Boston: Unwin Hyman, 1989.
- Liaw, A., and M. Wiener. 'Classification and Regression by randomForest'. *R News* 2, no. 3 (2002): 18–22.
- Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press, 2003.
- Medvedev, Dmitry. 'Vstrecha s voennosluzhashimi Vooruzhennykh Sil Rossii [Meeting with servicemen of the Armed Forces of Russia]'. *President.ru*, 14 August 2008, <http://kremlin.ru/text/appears/2008/08/205305.shtml>.
- Miller, Steven E., and Dmitry Trenin. *Vooruzhennye sily Rossii: vlast' i politika [The Russian Armed Forces: Power and Politics]*. Cambridge, MA: MIT Press, 2005.
- Monroe, Burt, Michael Colaresi and Kevin Quinn. 'Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict'. *Political Analysis* (forthcoming 2009).
- Mueller, John. *War, Presidents and Public Opinion*. New York: John Wiley & Sons, 1973.
- Ostrom, Charles W., and Brian L. Job. 'The President and the Political Use of Force'. *American Political Science Review* 80, no. 2 (1986): 541–66.
- Polikar, Robi. 'Ensemble Based Systems in Decision Making'. *IEEE Circuits and Systems Magazine* (2006): 21–45.
- Polsby, Nelson. *Congress and the Presidency*. Englewood Cliffs: Prentice-Hall, 1964.
- Poole, Keith. *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press, 2005.
- Quinn, Kevin, Burt L. Monroe, Michael Colaresi, Michael Crespin, and Dragomir Radev. 'How To Analyze Political Attention With Minimal Assumptions And Costs'. Society for Political Methodology, Davis, CA, 2006.
- Russett, Bruce. 'Economic Decline, Electoral Pressure, and the Initiation of Interstate Conflict'. In *Prisoners of War? Nation-States in the Modern Era*, ed. Charles Gochman and Alan Ned Sabrosky. Lexington, MA: D.C. Heath, 1989.
- Schliep, Klaus, and Klaus Hechenbichler. *kkn: Weighted k-Nearest Neighbors*. R package version 1.0–6 (2008).
- Scobell, Andrew. 'Soldiers, Statesmen, Strategic Culture and China's 1950 Intervention in Korea'. *Journal of Contemporary China* 8, no. 22 (1999): 477–97.
- Scobell, Andrew. 'Show of Force: Chinese Soldiers, Statesmen, and the 1995–1996 Taiwan Strait Crisis'. *Political Science Quarterly* 115, no. 2 (2000).
- Sechser, Todd. 'Are Soldiers Less War-Prone than Statesmen?' *Journal of Conflict Resolution* 48, no. 5 (2004): 746–74.
- Selivanova, Irina. 'Trans-Dniestria'. In *US and Russian Policymaking With Respect to the Use of Force*, ed. Jeremy R. Azrael and Emil A. Payin. Santa Monica, CA: RAND, 1996.

- Shulman, Stuart. 'Editor's Introduction'. *Journal of Information Technology and Politics* 5, no. 4 (2008): 353–4.
- Simmel, Georg. 'The Persistence of Social Groups'. *The American Journal of Sociology* 3, no. 5 (1898): 662–98.
- Skocpol, Theda. *States and Social Revolutions*. London: Cambridge University Press, 1979.
- Snowball Stemmer. Software. <http://snowball.tartarus.org/>
- Snyder, Jack. *The Ideology of the Offensive: Military Decision Making and the Disasters of 1914*. Ithaca, NY: Cornell University Press, 1984.
- Stewart, Philip D. 'Soviet Interest Groups and the Policy Process: The Repeal of Production Education'. *World Politics* 22, no. 1 (1969): 29–50.
- Stewart, Philip D., James W. Warhola, and Roger A. Blough. 'Issue Salience and Foreign Policy Role Specialization in the Soviet Politburo of the 1970s'. *American Journal of Political Science* 28, no. 1 (1984): 1–22.
- Vagts, Alfred. *A History of Militarism*. New York: Norton, 1937.
- Van Evera, Stephen W. *Causes of War: Power and the Roots of Conflict*. Ithaca, NY: Cornell University Press, 1999.
- Wallerstein, Immanuel. 'Third World Conflict in Soviet Military Thought: Does "New Thinking" Grow Prematurely Grey?' *World Politics* 42, no. 1 (1989): 31–63.
- Waltz, Kenneth. 'Electoral Punishment and Foreign Policy Crisis'. In *Domestic Sources of Foreign Policy*, ed. James Rosenau. New York: The Free Press, 1967.
- Wilkenfeld, Jonathan. 'Domestic and Foreign Conflict Behavior of Nations'. *Journal of Peace Research* 1 (1968): 56–69.
- Wilkenfeld, Jonathan, ed. *Conflict Behavior and Linkage Politics*. New York: David McKay, 1973.
- Yang, Y., and X. Liu. 'A Re-examination of Text Categorization Methods'. *Proceedings of SIGR* (November 1999).
- Zimmerman, William. *The Russian People and Foreign Policy: Elite and Mass Perspectives, 1993–2000*. Princeton: Princeton University Press, 2002.
- Zimmerman, William. 'Slavophiles and Westernizers Redux: Contemporary Russian Elite Perspectives'. *Post-Soviet Affairs* 11, no. 3 (2005): 183–209.
- Zimmerman, William, and Glenn Palmer. 'Words and Deeds in Soviet Foreign Policy: The Case of Soviet Military Expenditures'. *American Political Science Review* 77, no. 2 (1983): 358–67.
- Zinnes, Dina A., and Jonathan Wilkenfeld. 'An Analysis of Foreign Conflict Behavior of Nations'. In *Comparative Foreign Policy: Theoretical Essays*, ed. W.F. Hanrieder. New York: David McKay, 1971.

APPENDIX:

“USE OF FORCE AND RUSSIAN CIVIL-MILITARY RELATIONS: AN AUTOMATED CONTENT ANALYSIS”

Brandon M. Stewart and Yuri M. Zhukov
1 April 2009

CONTENTS

Contents.....	1
Appendix A: Sources for Russian documents	2
Russian Sources used (East-view database).....	2
Military and Security Periodicals.....	2
Government Publications.....	2
Central Newspapers	2
Appendix B: Codebook.....	4
Appendix C: Document Analysis	11
C.1: Text Processing	11
Stop-words:.....	12
Stop-stems:	12
C.2: Feature Selection	13
C.3 Unsupervised Model.....	14
C.4 Supervised Models.....	15
K-Nearest Neighbor.....	16
Adaboost.M1	16
Support Vector Machine	16
Random Forest	16
Uncertainty	17
Appendix D: Uncertainty	18
References	19

APPENDIX A: SOURCES FOR RUSSIAN DOCUMENTS

RUSSIAN SOURCES USED (EAST-VIEW DATABASE)

Military and Security Periodicals

- *Agentstvo voennykh novostei*
- *Armeiskii sbornik*
- *Boevaia vakhta*
- *Flag Rodiny*
- *Krasnaia zvezda*
- *Krasnyi voïn*
- *Morskaia gazeta*
- *Morskoi sbornik*
- *Na boevom postu*
- *Na strazhe Rodiny*
- *Nezavisimoe voennoe obozrenie*
- *Novosti razvedki i kontrrazvedki*
- *Orientir*
- *Rossiiskoe voennoe obozrenie*
- *Shchit i mekh*
- *Soldat Otechestva*
- *Soldat Rossii*
- *Strazh Baltiki*
- *Tekhnika i vooruzhenie*
- *Ural'skie voennye vesti*
- *VPK. Voenno-promyshlennyi kur'er*
- *VVS segodnia*
- *Vestnik voennoi informatsii*
- *Voennaia mysl'*
- *Voenno-istoricheskii zhurnal*
- *Voenno-meditsinskii zhurnal*
- *Voennye znaniia*
- *Voennyi diplomat*
- *Voennyi vestnik Iuga Rossii*
- *Voin Rossii*
- *Voprosy bezopasnosti*
- *Zakavkazskie voennye vedomosti*
- *Zarubezhnoe voennoe obozrenie*
- *Zashchita i bezopasnost'*

Government Publications

- *Diplomaticheskii vestnik*
- *Gosudarstvennaia Duma. Dnevnik zasedanii*
- *Gosudarstvennaia Duma. Parlamentskie slushaniia*
- *Gosudarstvennaia Duma. Povestka zasedanii Soveta Dumy*
- *Gosudarstvennaia Duma. Stenogramma zasedanii*
- *Parlamentskaia gazeta*
- *Prezident. Poslaniia*
- *Prezident. Vystupleniia*
- *Rossiiskaia gazeta*
- *Sovet Federatsii. Dnevnik zasedanii*
- *Sovet Federatsii. Obrashcheniia*
- *Sovet Federatsii. Stenogrammy zasedanii*
- *Sovet Federatsii. Zaiavleniia*

Central Newspapers

- *Argumenty i fakty*
- *Ekspert*
- *Ezhenedel'nyi zhurnal*
- *Finansovye Izvestiia*
- *Gazeta*
- *ITAR-TASS Daily*
- *InterFax-Vremia*
- *Itogi*
- *Izvestiia*
- *Kommersant. Daily*
- *Kommersant. Vlast'*
- *Komsomol'skaia pravda*
- *Konservator*
- *Krasnaia zvezda*
- *Kul'tura*

Central Newspapers (cont'd)

- *Literaturnaia gazeta*
- *Moskovskaia pravda*
- *Moskovskie novosti*
- *Moskovskii komsomolets*
- *Nasha versiia*
- *New Times*
- *Nezavisimaia gazeta*
- *Novaia gazeta*
- *Novoe vremia*
- *Novye Izvestiia*
- *Ogonek*
- *Pravda*
- *Profil'*
- *Rossiia*
- *Rossiiskaia gazeta*
- *Rossiiskie vesti*
- *Russkii Telegraf*
- *Russkii kur'er*
- *Sankt-Peterburgskie vedomosti*
- *Segodnia*
- *Slovo*
- *Sovetskaia Rossiia*
- *Tribuna*
- *Trud*
- *Vecherniaia Moskva*
- *Vedomosti*
- *Vek Versiia*
- *Vremia MN*
- *Vremia novostei*
- *Zavtra*

APPENDIX B: CODEBOOK

Unit of Analysis: Each observation represents a speech, press briefing, published article, interview or other public statement made by a political or military public figure in Russia. The original dataset included 7,920 observations, 5,143 of which remained in the cross-sectional dataset after the extraction of missing values (“NA” or “other topic/classification”) for the dependent variables **TOPIC** and **CLASS**. The period of observation is 12 February 1998 to 31 October 2008, which represent the release dates of the oldest and most recent statements in the collection.

This dataset is used for Models 1-3.

DOCUMENT_ID

Unique code assigned to each observation

DATE

Date of public statement’s release. Format: YYYYMMDD

NAME

Surname of author of statement

TS.SELECTION

0=Not included in training set

1=Included in training set (random selection of 300 observations)

TS.CLASS

0=Conservative (training set data only)

1=Activist (training set data only)

99=Not applicable/None of the above (training set data only)

NA=Not included in training set

Statements consistent with a *conservative* outlook (TS.CLASS=0) include:

- Expressions of a preference toward the use of military instruments of power only as a last resort:
Example: “К военной мощи следует прибегать, когда возможности других средств исчерпаны [Military power should be resorted to when the capabilities of other instruments are exhausted]”
- Expressions of a preference toward interagency solutions to security problems:
Example: “Военная безопасность Российской Федерации должна достигаться всей деятельностью государства - как в военной сфере, так и в политико-дипломатической, экономической, культурной, социальной и других [The military security of the Russian Federation should be attained through government-wide action — in the military sphere, as well as political-diplomatic, economic, cultural, social and others].

- Explicit doubts about the efficacy of military solutions in addressing specific political problems.
Example: “С террором необходимо бороться адекватными, но конституционными способами [Terror should be fought with adequate, but constitutional means].”
- Support for multilateral solutions to political and security problems.
Example: “Решение этих вопросов, особенно решение о применении силы, должно проходить исключительно через Совет Безопасности Организации Объединенных Наций [Decisions on such questions, particularly on the use of force, should be made exclusively through the United Nations Security Council].”
- General support for limited ends and means in foreign and defense policy.
Example: “Требуется определенная умеренность в определении и отстаивании национальных интересов, чтобы жестко отстаивать только действительно жизненно важные из них [A certain restraint is necessary in the definition and defense of national interests, so as to stringently defend only those, which are truly fundamental].”

Statements consistent with an *activist* outlook (TS.CLASS=1) include:

- Explicit support for the use of military power in a specific scenario
Example: “Выполняя свой воинский долг, вы хорошо понимали, что вы – это, по сути, последняя надежда незащищенных людей [While performing your military duty, you understood well that you, in essence, were the last hope of a defenseless people].”
- Support for military as most effective instrument of power in a given scenario.
Example: “Мы часто получаем официальные отзывы представителей ООН и ОБСЕ, которые признают наших миротворцев главным гарантом мира и стабильности [We often receive official feedback from UN and OSCE representatives, who admit that our peacekeepers are main guarantors of peace and stability].”
- Interpretation of the use or show of force as an effective means to certain political ends.
Example: “Применение ВМФ в мирное и военное время ... существенно повысит степень защищенности и реализуемости военно- стратегических и экономических интересов России в Мировом океане, позволит полностью или в значительной мере решить многие социальные, экономические и экологические проблемы приморских регионов Российской Федерации, а также международных отношений и сотрудничества с соседними морскими государствами [The use of the Navy in peacetime and in war ... will significantly increase the level of defensibility and realizability of Russia’s military-strategic and economic interests in the World Ocean, will enable the solution – in full or in significant part — of many social, economic and ecological problems of littoral areas of the Russian Federation, as well as international relations and cooperation with neighboring coastal states].”
- Support for unilateral solutions to interventionist crises.
Example: “Россия исторически была и останется гарантом безопасности народов Кавказа [Russia historically has been and will remain the guarantor of security for the peoples of the Caucasus].”
- General preference for foreign policy outcomes that revise, rather than reinforce the status quo.
Example: “Такого положения, такой опасности Россия не может себе позволить [Russia cannot permit such a state of affairs, such a threat to exist].”

CLASS

0=Conservative

1=Activist

NA=Not Applicable

Values assigned by an ensemble of supervised document classifiers (K-Nearest Neighbor, Adaboost.M1, Random Forest and Support Vector Machine), as outlined in Research Design section. TS.CLASS served as training input for document classification.

TOPIC

0=Realpolitik

1=Interventionist

NA=Not Applicable

Values assigned by Expressed Agenda Model as described in Research Design section, according to categories outlined in Table 1.

MILT

0= Author of statement is a Political Elite

1= Author of statement is a Military Elite

Military Elites include active duty and retired officers of field grade and above, who at the time of the statement's publication occupied a position of formal authority or informal influence in the Ministry of Defence or an affiliated institute/agency. Individuals from the following offices are included in the Military Elite category: Office of the Minister of Defence (uniformed personnel only), General Staff, Service/Branch Headquarters, Territorial Commands, high-visibility unit commands (such as peacekeeping contingents), professional military education institutions, military research institutions and the Academy of Military Sciences.

Political Elites include civilians, who at the time of the statement's publication occupied a position of formal authority or informal influence in the Government of Russia or an affiliated institute/agency. Individuals from the following offices are included in the Political Elite category: Executive Office of the President, Government of Russia (Prime Minister's Office), Security Council, Ministry of Foreign Affairs, Federal Security Service, Foreign Intelligence Service, Federation Council, State Duma, and the Council on Foreign and Defense Policy. Only individuals whose policy portfolios include foreign and security affairs – such as senior officials in cabinet ministries or members of the relevant parliamentary committees – were included in this sample.

CABT

1= Author is member of Presidential Administration or Security Council

0= Otherwise

PMNT

1= Author is member of State Duma or Federation Council

0= Otherwise

INFORMAL

0=Formal Authority

1=Informal Influence

In the cases examined here, a value of 0 (“Formal Authority”) was assigned for public statements made by individuals who at the time of publications occupied senior positions in the following structures (in alphabetical order):

- Emergency And Civil Defense State Committee
- Federal Assembly of Russia, Federation Council, Committee on CIS Affairs
- Federal Assembly of Russia, Federation Council, Committee on Defence and Security
- Federal Assembly of Russia, Federation Council,
- Federal Assembly of Russia, State Duma, Committee on CIS Affairs
- Federal Assembly of Russia, State Duma, Committee on Defence
- Federal Assembly of Russia, State Duma, Committee on Federal Budget Appropriations for Defense and State Security
- Federal Assembly of Russia, State Duma, Committee on International Affairs
- Federal Assembly of Russia, State Duma, Committee on Security
- Federal Assembly of Russia, State Duma, Committee on Veterans Affairs
- Federal Assembly of Russia, State Duma, Leadership
- Federal Border Service
- Federal Security Service
- Foreign Intelligence Service
- Government of Russia (Prime Minister’s Office)
- Kremlin, Presidential Civil Service Directorate
- Kremlin, Presidential Directorate for Interregional Relations and Cultural Contacts with Foreign Countries
- Kremlin, Presidential Executive Office
- Kremlin, Presidential Foreign Policy Directorate
- Ministry of Defence, Accommodation and Amenity Service, Main Accommodation and Amenity Directorate
- Ministry of Defence, Accommodation and Amenity Service,
- Ministry of Defence, Air Force, 13th State Research Institute
- Ministry of Defence, Air Force, 16th Air Army
- Ministry of Defence, Air Force, 37th Air Army (SOF)
- Ministry of Defence, Air Force, 4th Air and AAD Army
- Ministry of Defence, Air Force, 4th Center for Combat and Flight Training
- Ministry of Defence, Air Force, 5th Air and AAD Army
- Ministry of Defence, Air Force, 61st Air Army (Transport)
- Ministry of Defence, Air Force, 6th Air and AAD Army
- Ministry of Defence, Air Force, Aeronautical Service
- Ministry of Defence, Air Force, Central AF Command Point
- Ministry of Defence, Air Force, Far Eastern Group
- Ministry of Defence, Air Force, Headquarters

- Ministry of Defence, Air Force, Military Academy of Aerospace Defence
- Ministry of Defence, Airborne Troops, Headquarters
- Ministry of Defence, General Staff
- Ministry of Defence, General Staff, Directorate of Communications
- Ministry of Defence, General Staff, Directorate of Military Topography
- Ministry of Defence, General Staff, Directorate of Radio-Electronic Warfare
- Ministry of Defence, General Staff, General Intelligence Directorate
- Ministry of Defence, General Staff, General Mobilization Directorate
- Ministry of Defence, General Staff, General Operational Directorate
- Ministry of Defence, General Staff, Military Academy of the General Staff
- Ministry of Defence, Ground Forces Command, Far Eastern Military District
- Ministry of Defence, Ground Forces Command, Headquarters
- Ministry of Defence, Ground Forces Command, Leningrad Military District
- Ministry of Defence, Ground Forces Command, Moscow Military District
- Ministry of Defence, Ground Forces Command, North Caucasus Military District
- Ministry of Defence, Ground Forces Command, Siberian Military District
- Ministry of Defence, Ground Forces Command, Volga-Ural Military District
- Ministry of Defence, Navy, Baltic Fleet
- Ministry of Defence, Navy, Black Sea Fleet
- Ministry of Defence, Navy, Caspian Flotilla
- Ministry of Defence, Navy, Headquarters
- Ministry of Defence, Navy, Northern Fleet
- Ministry of Defence, Navy, Pacific Fleet
- Ministry of Defence, Office of First Deputy Minister of Defence, Main Directorate for Combat Training and Service
- Ministry of Defence, Office of First Deputy Minister of Defence, Military Inspection
- Ministry of Defence, Office of First Deputy Minister of Defence,
- Ministry of Defence, Office of State Secretary, Directorate of State Civil Service
- Ministry of Defence, Office of State Secretary, Executive and Legislative Powers Cooperation Branch
- Ministry of Defence, Office of State Secretary, General Directorate for Morale
- Ministry of Defence, Office of State Secretary, Main Personnel Directorate
- Ministry of Defence, Office of State Secretary,
- Ministry of Defence, Office of the Deputy Minister,
- Ministry of Defence, Office of the Deputy Minister for Armament, Executive Office
- Ministry of Defence, Office of the Deputy Minister for Armament, General Rocket Artillery Directorate
- Ministry of Defence, Office of the Deputy Minister for Armament, General Tank-Automotive Directorate
- Ministry of Defence, Office of the Deputy Minister for Armament,
- Ministry of Defence, Office of the Deputy Minister for Financial and Economic Affairs, Civil Disbursements Directorate
- Ministry of Defence, Office of the Deputy Minister for Financial and Economic Affairs, General Finance Directorate
- Ministry of Defence, Office of the Deputy Minister for Financial and Economic Affairs
- Ministry of Defence, Office of the Deputy Minister for Logistics, Directorate of Logistics

- Ministry of Defence, Office of the Deputy Minister for Logistics, General Medical Directorate
- Ministry of Defence, Office of the Deputy Minister for Logistics, General Medical Directorate
- Ministry of Defence, Office of the Deputy Minister for Logistics, General Medical Directorate
- Ministry of Defence, Office of the Deputy Minister for Logistics
- Ministry of Defence, Office of the Minister of Defence, Administration of MoD Affairs
- Ministry of Defence, Office of the Minister of Defence, Directorate of Information and Public Relations
- Ministry of Defence, Office of the Minister of Defence, Expert Center
- Ministry of Defence, Office of the Minister of Defence, Financial Inspection
- Ministry of Defence, Office of the Minister of Defence, Main Directorate for International Cooperation
- Ministry of Defence, Office of the Minister of Defence, Main Legal Directorate
- Ministry of Defence, Office of the Minister of Defence, Secretariat
- Ministry of Defence, Office of the Minister of Defence
- Ministry of Defence, Railroad Troops, Headquarters
- Ministry of Defence, Space Forces, Headquarters
- Ministry of Defence, Strategic Rocket Forces, Headquarters
- Ministry of Finance
- Ministry of Foreign Affairs
- Ministry of Internal Affairs
- Ministry of Justice
- Office of the Prosecutor General
- Security Council
- State Narcotics Control Service

A value of 1 (“Informal Influence”) was assigned for public statements made by individuals who at the time of publications occupied senior positions in the following structures (in alphabetical order):

- Academy of Military Sciences
- Russian Academy of Sciences
- Council on Foreign and Defence Policy

PFSCR

Composite Press Freedom Score (0-100 ordinal scale). Measured annually.

Source: “Freedom of the Press Historical Data,” Freedom House,

<http://www.freedomhouse.org/template.cfm?page=274>;

“Freedom of the Press 2007 Survey,” Freedom House,

<http://www.freedomhouse.org/template.cfm?page=107&year=2007>;

“Freedom of the Press 2008 Survey,” Freedom House,

<http://www.freedomhouse.org/template.cfm?page=362>.

Coding specifications available at:

<http://www.freedomhouse.org/uploads/fopo8/Methodology2008.pdf>

W1, W2, W3

Three vectors of probabilities of observing Class 1 (Conservative), Class 2 (Activist) or Class 3 (Not Applicable) for each document. These probabilities are obtained from the normalized ensemble classifier weights and can be used to simulate the imputation draws for uncertainty analyses.

DISAGREE

Mean level of disagreement between military and political elites in a given month.

Measured as Cartesian distance between mean monthly military and political elite classifications for CLASS and TOPIC:

$$\sqrt{[(\text{MIL.CLASS}_t - \text{POL.CLASS}_t)^2 + (\text{MIL.TOPIC}_t - \text{POL.TOPIC}_t)^2]}$$

SC.MILP

Proportion of Security Council officials with professional military background.

$$\text{SC.MIL} / \text{SC.TOT}$$

where SC.MIL is the number of members of the Russian Security Council with a professional military background (defined as attendance of a professional military education institution) and SC.TOT is the number of seats on the Security Council.

Source: Security Council of the Russian Federation,
<http://www.scrf.gov.ru/persons/sections/parent/>

APPENDIX C: DOCUMENT ANALYSIS

This appendix includes supplemental replication details. It generally does not repeat information found in the body of the text, but focuses on additional replication parameters. Full replication code and data will be made available upon request.

Document Analysis Summary

Task	Use of Force	Topic Model
<i>Obtain Source Material</i>	~8000 Press Releases and Interviews (see Appendix D)	
<i>Document Preprocessing</i>	Russian Stemming Feature Extraction	
<i>Feature Selection</i>	Latent Semantic Analysis	Latent Dirichlet Allocation
<i>Pre-Analysis Human Intervention</i>	300 Document Set	# of Topics Set
<i>Methods</i>	Ensemble Supervised Learning	Unsupervised Expressed Agenda Model
<i>Post-Processing</i>	Distance Analysis	Topic Interpretation Distance Analysis

C.1: TEXT PROCESSING

We made the decision to analyze texts in their native Russian. Machine translation is notoriously unreliable, despite the relatively relaxed assumptions of bag-of-words analysis. In order to provide the best results, words must be consistent in their meaning across texts. This requires that multiple words meaning different things in Russian not be translated to the same word in English. Unfortunately, machine translation is not even up to simple word-for-word translation that ignores lexical ordering. To use an example from the training set, the sentence “применение ВМФ в мирное и военное время ... существенно повысит степень защищенности ... интересов России” should translate to “employment of the Navy in peacetime and wartime ... would substantially increase the level of *defensibility* ... of Russia’s interests.” Meanwhile, machine translation (BabelFish) translates the sentence as “the application of the Navies in the peaceful and military time... will significantly increase the *vulnerability* of the ... interests of Russia.” Thus, the machine-translated text would receive a coding of “Conservative,” since it would appear to argue that force employment makes Russia *less* safe, while the original Russian clearly argues the opposite.

Text processing was performed using a custom Python script written under Python 2.7 to handle multiple encoding formats of the Cyrillic text. The python script was built using the BeautifulSoup HTML parsing library and the Snowball Project Russian stemmer. It creates a *tf*idf* matrix and a count matrix, which include only unigrams which appear in more than 1% of the text and less than

99% of the texts. A list of stop-words and stop-stems were removed from the analysis. The stop-words were generated from a list provided with the snowball stemmer. The stop-stems were generated empirically by the authors in the process of refining the Expressed Agenda model.

Stop-words:

и,в,во,не,что,он,на,я,с,со,как,а,то,все,она,так,его,но,да,ты,к,у,же,вы,за,бы,по,только,ее,мне,было,вот,от,меня,еще,нет,о,из,е
му,теперь,когда,даже,ну,вдруг,ли,если,уже,или,ни,быть,был,него,до,вас,нибудь,опять,уж,вам,сказал,ведь,там,потом,себя,ниче
го,ей,может,они,тут,где,есть,надо,ней,для,мы,тебя,их,чем,была,сам,чтоб,без,будто,человек,чего,раз,тоже,себе,под,жизнь,будет
,ж,тогда,кто,этот,говорил,того,потому,этого,какой,совсем,ним,здесь,этом,один,почти,мой,тем,чтобы,нее,кажется,сейчас,были,
куда,зачем,сказать,всех,никогда,сегодня,можно,при,наконец,два,об,другой,хоть,после,над,больше,тот,через,эти,нас,про,всего,
них,какая,много,разве,сказала,три,эту,моя,впрочем,хорошо,свою,этой,перед,иногда,лучше,чуть,том,нельзя,такой,им,более,вс
егда,конечно,всю,между,меня,мне,мною,мною,ты,тебя,тебе,тобой,тобою,он,его,ему,им,него,нему,ним,она,ее,эи,ею,нее,нэи,не
ю,оно,его,ему,им,него,нему,ним,мы,нас,нам,нами,вы,вас,вам,вами,они,их,им,ими,них,ним,ними,себя,себе,собой,собою,этом,э
та,это,эти,этого,эты,это,эти,этого,этой,этого,этих,этому,этой,этому,этим,этим,этой,этим,этой,этими,этом,этой,этом,этих,тот,т
а,то,те,того,ту,то,те,того,той,того,тех,тому,той,тому,тем,тем,той,тем,тою,теми,том,той,том,тех,весь,весь,вся,все,все,всего,всю,вс
е,все,всего,всей,всего,всех,всему,всей,всему,всем,всем,всей,всем,всего,всеми,всем,всей,всем,всех,сам,сама,само,сами,самого,са
му,само,самых,самого,самой,самого,самых,самому,самой,самому,самим,самим,самой,самим,самоу,самими,самом,самой,само
м,самых,быть,бы,буд,быв,есть,суть,име,дел,мог,мож,мочь,уме,хоч,хот,долж,можн,нужн,нельзя,Будут,люб,рич,ита,ладимир,на
де,письм,нын,де,гост,пасиб,даст,ладимир,каков,арт,полн,Оссийск,каков,яв,видет,ладимир,будут,люб,рич,ос,арт,будут,люб,ри
ч,будут,люб,рич,де,пасиб,ладимир,нын,нужд,рич,ита,яв,полн,лед,оссийск,будут,люб,рич,видет,Руз,важа,проч,будут,люб,рич,я
в,едерац,будут,люб,рич,горазд,наде,одн,де,дин,ладимир,будут,люб,рич,ожн,рич,люб,будут,проч,знал,будут,рми,ладимирович,
каков,видим,видим,ладимирович,рми,давл,рми,важа,леж,случ,знан,ладимирович,сег,рамк,рми,давл,ладимирович,рми,един,
оп,случ,знан,ладимирович,видим,рми,дат,видим,знан,леж,ладимирович,ладимирович,видим,дат,заметн,ладимирович,онечн,
сто,отмеч,личн,реч,основан,господин,парт,должн,видел,кром,ольк,рот,тат,кром,дава,вообщ,уш,трет,ком,реч,п,идт,ез,каза,пят
,сет,ком,иха,ед,двойн,господ,суд,мор,сет,руг,понятн,ред,видел,ита,ез,пят,шест,двойн,ред,ольк,сто,ед,отмеч,личн,реч,тип,онеч
н,основн,ведет,счет,месяц,собра,уш,трет,ез,каза,сет,тат,кром,дава,вообщ,сет,круг,п

Stop-stems:

будут,люб,рич,ита,ладимир,наде,письм,нын,де,гост,пасиб,даст,ладимир,каков,арт,полн,Оссийск,каков,яв,видет,ладимир,б
удут,люб,рич,ос,арт,будут,люб,рич,будут,люб,рич,де,пасиб,ладимир,нын,нужд,рич,ита,яв,полн,лед,оссийск,будут,люб,рич,вид
ет,Руз,важа,проч,будут,люб,рич,яв,едерац,будут,люб,рич,горазд,наде,одн,де,дин,ладимир,будут,люб,рич,ожн,рич,люб,будут,п
роч,знал,будут,рми,ладимирович,каков,видим,видим,ладимирович,рми,давл,рми,важа,леж,случ,знан,ладимирович,сег,рамк,
рми,давл,ладимирович,рми,един,оп,случ,знан,ладимирович,видим,рми,дат,видим,знан,леж,ладимирович,ладимирович,види
м,дат,заметн,ладимирович,онечн,сто,отмеч,личн,реч,основан,господин,парт,должн,видел,кром,ольк,рот,тат,кром,дава,вообщ
,уш,трет,ком,реч,п,идт,ез,каза,пят,сет,ком,иха,ед,двойн,господ,суд,мор,сет,руг,понятн,ред,видел,ита,ез,пят,шест,двойн,ред,ол
ьк,сто,ед,отмеч,личн,реч,тип,онечн,основн,ведет,счет,месяц,собра,уш,трет,ез,каза,сет,тат,кром,дава,вообщ,сет,круг,п,военизд
ат,ефимович,пыт,седьм,политолог,гот,миров,воз,едв,пыт,восем,явл

A small number of documents were removed because they contained no word counts. The Mutual Information Criterion was implemented separately in R.

C.2: FEATURE SELECTION

After removing stems and creating the document-by-term matrix $\mathbf{X}$ ($d \times t$), we begin by weighting the terms for maximum effect. We use two different criteria: the widely used Term Frequency by Inverse Document Frequency ($tf \cdot idf$) and a Mutual Information Criterion. Both criteria are designed to weight words more highly when they are relatively rare in the corpus as a whole. $tf \cdot idf$ is a standard in the computational linguistics literature (Manning & Schütze 2003) and has been defended on theoretical grounds (Papineni 2001). The mutual information criterion is advocated by Hillard(2003) and includes a normalizing element. The formulas for each are as follows:

$$tf * idf = \mathbf{X}_{ij} * \log \frac{d}{d_{j>1}}$$

$$\text{mutinf}_{ij} = \log \frac{p(w, t)}{p(w)p(t)} = \log \frac{p(w|t)}{p(w)}$$

where documents are indexed by $\mathbf{i} = 1 \dots d$ Docs and terms are indexed by $\mathbf{j} = 1 \dots t$ terms. After weighting the term counts we still have a term matrix that has dimensions ($d \times t$). For dimensionality reduction we employ Latent Semantic Analysis (LSA). LSA was explored for document indexing as a way of reducing the feature space and linking words that coexist in the document as a whole (Deerwester et. al 1990), and has become a standard tool of topic models in the computational linguistics community (Manning & Schütze 2003, Manning et. al 2008, Tzoukerman et al. 2003).

LSA uses singular value decomposition (SVD) to map the weighted count matrix into a lower dimensional space by combining associated words into single features.¹ This feature reduction has the added benefit of addressing issues of synonymy. LSA allows two documents which may not share a word in common to be mapped onto a similar feature space because the words coexist in the corpus as a whole. For example, while one document may have the word “talk” and the other has the word “speak”, LSA would most likely group them into the same feature space. We used two separate matrix reductions, one where we reduced to 100 features, and the other where we reduced to 500 features.

¹ This is similar to Principle Component Analysis (PCA) but does not require a singular matrix. Using SVD on the transpose of the weighted count matrix $\mathbf{X}^T (t \times d)$ yields matrices $\mathbf{U} (t \times m)$, $\mathbf{D} (m \times m)$ and $\mathbf{V}^T (m \times d)$, where $\mathbf{D}$ is a diagonal matrix whose singular values descend. By selecting the top k values, where $k < m$ and zeroing out the other values to create $\mathbf{D}^*$, $\mathbf{U}\mathbf{D}^*\mathbf{V}^T$ becomes the least squares approximation of $\mathbf{X}^T$. With that intuition, we can see that $\mathbf{D}\mathbf{V}^T$ is the matrix $\mathbf{X}^T$ projected into a lower dimensional space. By taking the transpose, we now have a matrix $\mathbf{X}^* (d \times f)$ where f is a number of features which is less than t . Similar feature reduction is completed in the Expressed Agenda Model using Latent Dirichlet Allocation, which is analogous to (although not the same as) probabilistic LSA. Notation conventions for the decomposition matrices follow the R notation for Singular Value Decomposition.

The described Latent Semantic Analysis was generated through use of the Singular Value Decomposition routine in the **base** R package. The *tf*idf* matrix was reduced to the top $k=100$ values. The Mutual Information matrix was reduced to the top $k=500$ values. These values were chosen empirically by using cross-validation to check accuracy.

Table 1: Stems by Topic

	Russian Stems ²	Translations	Category
Topic 1	Россия страст[ь] свободн[о] сыгра[ть] полиц[ия] полицейск[ие] готов[ность] этническ[ие] кпрф межнациональн[ый]	Russia passion free play police (noun) police (adj.) readiness ethnic CPRF (Communists) interethnic	<i>Interventionist</i>
Topic 2	войт[и] армен[ия] годовщин[а] военнослужа[щие] войсков боготов[ность] [за]служен[ный] официальн[ый] частн[ый] возведен[ный]	enter Armenia anniversary military servicemen forces military readiness deserved official private elevated	<i>Realpolitik</i>
Topic 3	боготов[ность] срочн[о] войт[и] действительн[о] управля[ть] противовес задел сильн[ый] опережа[ть] систематическ[и]	military readiness immediately enter really control counterweight backup strong surpass systematically	<i>Realpolitik</i>
Topic 4	военнослужа[щие] воплот сильн[ый] находя[сь] готов[ность] безработиц[а] джордж науч[ное] Войсков москв[а]	military servicemen bulwark strong be situated readiness unemployment George scientific forces Moscow	<i>Realpolitik</i>

C.3 UNSUPERVISED MODEL

The Expressed Agenda model was provided by Justin Grimmer. Upon publication of Grimmer (2009) the package will be released for R under the CRAN server. In order to determine the appropriate number of clusters we iterated over various values of k from 3 clusters to 15. We

² Prefixes and suffixes and translated words are illustrative and reflect one potential variation for “full” words expanded from the stems. In most cases, multiple variations are possible.

finally settled on $k=6$ at which point documents were labeled according to the top 20 words in the mutual information matrix. Table 1 provides the top 10 mutual information stems for the four labeled topics. The other two clusters were dropped as irrelevant (coded NA) and were removed from the model.

C.4 SUPERVISED MODELS

We employ four different machine learning approaches in our ensemble classifier, two of which are ensemble classifiers in their own right: K-Nearest Neighbor, Adaboost.M1 algorithm, Random Forest and Support Vector Machine. Each classifier is discussed in turn and Figure 1 shows the entire system.

K-Nearest Neighbor proceeds from an intuitive assumption that an unknown case can be classified according to its k nearest neighbors in the parameter space (in this case using Minkowski distance). We selected $k=5$ and chose an un-weighted version of the model, where all five neighbors were treated equally (Hechenbiechler & Schliep 2004).

The Adaboost.M1 algorithm is a relatively recent discovery in machine learning, generating considerable interest following its introduction in the late 1990s (Polikar 2006). Adaboost.M1 uses several instances of a WeakLearner (in this implementation a classification tree) to generate hypotheses using data randomly drawn from the training distribution. The distribution is then iteratively updated to include instances misclassified by the first algorithm. This proceeds until a weighted majority vote occurs, which yields the final classification. Freund and Schapire (1997) provide a theoretical defense of the empirical success of the classifier, showing that as long as the error is less than .5 on each instance, the total error is bounded at the top and should decrease asymptotically with each iteration.

The Random Forest is similar to the Adaboost algorithm but uses a different approach. It is also an ensemble technique and uses classification trees as its sub-component. Rather than iteratively training near examples missed by the classifier previously, the trees are grown using bootstrapped versions of the data and by choosing k nodes for which to search for a split. This introduces random perturbations into the data which generate different results in each tree and prevents over-fitting, a common problem with decision trees (Breiman 1999, Breiman 2001a, Breiman 2001b).

Support Vector Machine (SVM) is easily the most popular machine learning algorithm in political science due to its easy implementation and broad utility compared to other techniques (Yang & Liu 1999). SVM fits a hyper-plane to the feature space, which separates two categories of points from each other and maximizes the marginal distance between the nearest points and the surface. A cost function determines the penalization of the soft-margins when inevitably all the points don't fall on one side of the plane. The logistical difficulty with SVM is that it is designed for dichotomous classification. We deal with this issue by using a common solution, running the SVM three times with each classification pitted against both others (NA versus ["Activist" + "Conservative"], "Activist" versus ["Conservative" + NA] etc.) and the classification with the greatest marginal distance from the hyper-plane is deemed the classifier's choice.

The first three algorithms are trained on the LSA-reduced $tf*idf$ feature matrix ($d \times 100$), and the SVM is trained on the LSA-reduced mutual information feature matrix ($d \times 500$). The

difference in feature sets and algorithms provides a diverse approach, beneficial given that no single classifier is superior for all classification problems.

The code for model estimation in each supervised learning technique is given below. Reference to the R package help file will illuminate the significance of chosen parameters and defaults:³

K-Nearest Neighbor

```
results <- kkn(TRAININGSET_VALUE ~., train = x, test = xtest, k = 5, kernel =
  "rectangular")
```

Adaboost.M1

```
results <- adaboost.M1(TRAININGSET_VALUE ~ ., data = x[-sample,], boos = TRUE, mfinal =
  100, coeflearn = 'Breiman', minsplit = 5, cp = 0.01, maxdepth = 18)
```

Support Vector Machine

```
results <- svm(path(x,NAvalue, epsilon=1e-6, lambda=0.1)
  predNA <- predict(results, xtest, lambda=1)
results <- svm(path(x,Avalue, epsilon=1e-6, lambda=0.1)
  predA <- predict(results, xtest, lambda=1)
results <- svm(path(x,Cvalue, epsilon=1e-6, lambda=0.1)
  predC <- predict(results, xtest, lambda=1)
```

Random Forest

```
results <- randomForest(x=x, y=y)
RFpredictions <- predict(results, data, type="response")
```

In order to effectively weight the predictions from the classifiers, we obtain accuracy measures by cross-validation for each algorithm. Using our set of 300 coded documents, we randomly sample 275, train the algorithm and then test on the out of sample 25 documents. We repeat this simulation 10,000 times for each algorithm to attain out-of-sample accuracy results.

We then weight the classifiers predictions according to the following formula where p is the number of correct out of sample predictions divided by the number of total out of sample predictions:

$$w_{class} = \log \frac{p}{1 - p}$$

Given these weights, we simulate the training of each classifier on the same sample of 275, apply weights to the predictions and attain a final prediction and then verify against the out of sample 25. We repeat this process 10,000 times to produce the accuracy rates for the ensemble system as a whole. Accuracies for the individual classifiers and overall are given in the table below. The

³ Note that the full R-Code and replication instructions are available in Appendix H. Code is provided here mainly as a reference for the critical specifications.

accuracy of the system is lower than any individual classifier, but not dramatically. Nonetheless, the ensemble approach guards against the idiosyncrasies of any given classifier.

Accuracy Rates

Classifier	Out-Of-Sample Error Rates
K-Nearest Neighbor	37.81%
Adaboost M.1	35.2%
Random Forest	34.36%
Support Vector Machine	48.08%
Ensemble	33.97%

Uncertainty

We use the estimated weights to incorporate the uncertainty of the classification procedure into our analysis, a form of uncertainty which is often discarded. We infer probabilities of categorization from the individual weights using the following formula:

$$p(Cat_i) = \frac{w_1 * Class1_i + w_2 * Class2_i + w_3 * Class3_i + w_4 * Class4_i}{w_1 + w_2 + w_3 + w_4}$$

where $ClassN_i$ is the dichotomous prediction of Classifier N for category i . We then redraw the data 10,000 times using a distribution defined by the probabilities for each class and observation. We then repeat the models using the new draws of the data. The mean of the coefficients is used as the coefficient of the estimate, and the standard deviation of the coefficients is used as the standard error, which incorporates the uncertainty in the classification process.

The probabilities of accuracy are measured as the probability of a correct classification on 10,000 cross-validation runs training on a random selection of 275 documents and holding back 25 documents. Probabilities are based only on this collection of 25 out-of-sample classification rates.

Estimates on replication will be imprecise due to the use of simulation in the text. However due to the large number of cross-validation runs, the estimates should converge. The process is extremely computationally intensive and over 10,000 runs the code should be expected to take several hours to run even on a newer machine.

APPENDIX D: UNCERTAINTY

Whether in automated or manual document analysis, error rates or inter-coder reliability rates are rarely taken into account at the level of estimation. Hopkins and King (2008) present a supervised document classification model called ReadMe that helps to ameliorate this deficiency by removing the error prone step of individual document classification. The algorithm, in contrast to supervised methods in the computer science literature, computes the distribution of document classifications rather than to classify each individual document. Unfortunately, we need individual level data, both for the analysis of our first three hypotheses and because several of the months in the time series have too few documents for an adequate distribution.

We can understand this procedure in two ways which are theoretically distinct but in practice the same. In the first understanding, the entire DV can be treated as missing, with each sample being a multiple imputation estimate (King et al. 2000). The second understanding is that we are performing a weighting of observations commensurate with our understood uncertainty. In either case, the procedure is the same. The one important difference is that under the construct of multiple imputation, we can include all of our meta-data into the classification process (the author, affiliation, date etc.); to fail to include any data that will eventually be used in the analysis from the imputation process can result in bias. We tested the inclusion of the meta-data and there was no difference in the analyses.

REFERENCES

- Alfaro, E., Gamez, M. and Garcia, N. 2007: "Multiclass corporate failure prediction by Adaboost.M1." *International Advances in Economic Research*, Vol 13, 3, pp. 301–312.
- Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*. 3: 993-1022.
- Blei, David and John Lafferty. 2006. "Dynamic Topic Models." *Proceedings of the 23rd International Conference on Machine Learning* 23.
- Breiman, Leo. 2001a, "Random Forests." *Machine Learning*. 45(1), 5-32.
- Breiman, Leo. 2001b. "Statistical Modeling: The Two Cultures." *Statistical Science*. 16:3, 199-231.
- Breiman, Leo. 2002, "Manual On Setting Up, Using, And Understanding Random Forests V3.1", Research Note.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer and Richard Harshman. 1990. "Indexing by latent semantic analysis." *Journal of the American Society for Information Science*, 41(6), 391-407.
- Embretson, Susan E. and Steven P. Reise. 2000. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Freund, Y. and Schapire, R.E. 1996: "Experiments with a New Boosting Algorithm". In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156, Morgan Kaufmann.
- Fox, John. 2008. car: Companion to Applied Regression. R package version 1.2-9.
- Freund, Y. and Schapire, R.E. 1997. "Decision-theoretic generalization of on-line learning and an application to boosting" *Journal of Computer and System Sciences*, vol. 55 no. 1, pp. 119-139.
- Gelman, Andrew, Gary King and Chuanhai Liu. 1999. "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys." *Journal of the American Statistical Association*, Vol. 93, No. 443: Pp. 846-857.
- Godbole, Namrata, Manjunath Srinivasaiah and Steven Skiena. 2007. "Large-Scale Sentiment Analysis for News and Blogs." *Proceedings of International Conference on Weblogs and Social Media*.
- Grimmer, Justin. 2008. "A Bayesian Hierarchical Topic Model for Political Texts: Supplemental Appendix." Working Paper.
- Grimmer, Justin. 2009. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." Working Paper.
- Hastie, Trevor, Saharon Rosset, Robert Tibshirani and Ji Zhu. 2004. "The Entire Regularization Path for the Support Vector Machine." *Journal of Machine Learning Research*. Vol 5, 1391-1415.
- Trevor Hastie 2006. svmPath: svmPath: the SVM Path algorithm. R package version 0.92. <http://www-stat.stanford.edu/~hastie/Papers/svmPath.pdf>
- Hillard, Dustin. 2003. "Topic Classification for Conversational Speech using Support Vector Machine and Latent Semantic Analysis." Working Paper.
- Hillard, Dustin, Stephen Purpura, and John Wilkerson. 2007. "An Active Learning Framework for Classifying Political Text." Midwest Political Science Association 65th Annual National Conference.
- Hillard, Dustin, Stephen Purpura and John Wilkerson. 2008. "Computer Assisted Topic Classification for Mixed Methods Social Science Research." *Journal of Information Technology and Policy*. 4(4).
- Hopkins, Daniel and Gary King. 2008. "A Method of Automated Nonparametric Content Analysis for Social Science." Working Paper. Previously presented at Midwest Political Science Association and the Society for Political Methodology.
- Imai, Kosuke, Gary King, and Olivia Lau. 2007. "negbin: Negative Binomial Regression for Event Count Dependent Variables" in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," <http://gking.harvard.edu/zelig>
- Imai, Kosuke, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," <http://GKing.harvard.edu/zelig>.
- Imai, Kosuke, Gary King, and Olivia Lau. 2008. "Toward A Common Framework for Statistical Analysis and Development." *Journal of Computational and Graphical Statistics*, Vol. 17, No. 4 (December), pp. 892-913.

- Joachims, T. 1998. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of the European Conference on Machine Learning (ECML).
- Liaw, A. and M. Wiener. 2002. Classification and Regression by randomForest. *R News* 2(3), 18--22.
- Manning, Christopher D. and Hinrich Schütze. 2003. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.
- Manning, Christopher D., Prabhakar Raghavan and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Monroe, Burt, Michael Colaresi and Kevin Quinn. 2009. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." Forthcoming in *Political Analysis*.
- Papineni, Kishore. 2001. "Why inverse document frequency?" *NAACL Proceedings*, 25-32.
- Polikar, Robi. 2006. "Ensemble Based Systems in Decision Making." *IEEE Circuits and Systems Magazine*. 21-45.
- Porter, M. F. 1980. "An algorithm for suffix stripping." *Program* 14(3):130-137.
- Purpura, Stephen and Dustin Hillard. 2006. "Automated Classification of Congressional Legislation." The 7th Annual International Conference on Digital Government Research. May 21-24, San Diego, CA.
- Purpora, Stephen, Dustin Hillard and Philip Howard. 2006a. "A Comparative Study of Human Coding and Context Analysis against Support Vector Machines (SVM) to Differentiate Campaign Emails by Party and Issues." Working Paper.
- Quinn, Kevin, Burt L. Monroe, Michael Colaresi, Michael Crespin, and Dragomir Radev. 2006. "How To Analyze Political Attention With Minimal Assumptions And Costs." Society for Political Methodology, Davis, CA.
- Schliep, Klaus and Klaus Hechenbichler (2008). kkn: Weighted k-Nearest Neighbors. R package version 1.0-6.
- Schrodt, Philip, Palmer, Glenn and Hatipoglu, Mehmet. 2008. "Automated Detection of Militarized Interstate Disputes Using Document Classification Algorithms" Paper presented at the annual meeting of the APSA 2008 Annual Meeting, Hynes Convention Center, Boston, Massachusetts, Aug 28, 2008 Online <WEBMAIL/PDF>. 2009-02-16 <http://www.allacademic.com/meta/p278383_index.html>
- Shulman, Stuart. 2008. "Editor's Introduction." *Journal of Information Technology & Politics*. 5(4), 353-354.
- Snowball Stemmer. Software. <http://snowball.tartarus.org/>
- Tzoukerman, Evelyne, Judith Klavans and Tomek Strzalkowski. 2003. "Information Retrieval." *Oxford Handbook of Computational Linguistics*. Ed. Ruslan Mitkov.
- Vapnic, V. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.
- Venables, W. N. and Ripley, B. D. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.