AMERICAN SOCIOLOGICAL REVIEW

OFFICIAL JOURNAL OF THE AMERICAN SOCIOLOGICAL ASSOCIATION

ONLINE SUPPLEMENT
to article in
AMERICAN SOCIOLOGICAL REVIEW, 2019, VOL. 84

Rhetorics of Radicalism

Daniel Karell
New York University Abu Dhabi

Michael Freedman
Massachusetts Institute of Technology

Part A. METHODOLOGICAL DETAILS OF THE COMPUTATIONAL ABDUCTIVE ANALYSIS

A.1. Preprocessing of text data

To conduct our computational textual analysis of documents in multiple languages, we prepared the documents in a series of steps (Grimmer and Stewart 2013; Lucas et al. 2015). First, all publications were digitized and converted to plain text format. Then, the non-English publications were translated into English. Many of the originally non-English publications had already been translated by human translators, such as the entire Pashto collection from the Taliban Sources Project (Strick van Linschoten and Kuehn 2018). For the non-English documents not previously translated, we hired professional translators to translate a random selection. The selection included publications by *Jabha-yi Nijat-i Milli-yi Afghanistan* (from Dari), *Jam'iyyat-i Islami-yi* (from Arabic and Urdu), the Haqqani Organization (from Arabic), *Jami'ah al Da'wah ila al-Quir'an wa al-Sunnah* (from Arabic), Miramshah College (from Arabic), and *Maktab al-Khidamat* (from Arabic). The size of the selection was limited by monetary resources.

The remaining non-English documents in the corpus were translated using translateR[1] (Lucas et al. 2015). This process induced data loss, as is usually the case. We attempt to account for the differential rates of this data loss across our documents by including a covariate for language in our structural topic models, as explained in more detail below. We recommend that future research re-translate the computer translated documents, because artificial intelligence translation tools are constantly improving. For example, as this article was accepted for publication, Microsoft released version 3 of Translator Text API; we had used version 2. To support improvements to the existing automated translations, we offer access to the non-translated versions of our documents at https://osf.io/jx756/.

After translating the entire corpus to English, the text was preprocessed (Grimmer and Stewart 2013) using the standard settings in the structural topic model package, or stm,[2] for R (Roberts, Stewart, and Airoldi 2016). This included splitting text strings into individual terms, converting all characters to lower case, stemming the terms, and removing special characters, numbers, stop-words, and infrequent terms.

These preprocessing changes to the words in the corpus are standard, but, as Denny and Spirling (2018) point out, each one can influence the results. To assess the impact of preprocessing on our analysis, we used the preText[3] R package (Denny and Spirling 2018). This package allowed us to estimate the effect of each preprocessing step on a "preText score." Negative coefficients mean the step reduces the "unusualness" of results for our corpus; positive coefficients indicate that conducting the preprocessing will likely lead to more "unusual" results.

Our preText analysis suggests that the preprocessing steps were overall unlikely to produce unusual findings (Figure S1). Four of the six steps were estimated to either have no effect or to produce more normal results. We chose to include the two steps that were likely to increase unusual findings (i.e., removing punctuation and numbers) because we did not want these kinds of characters to influence the findings.

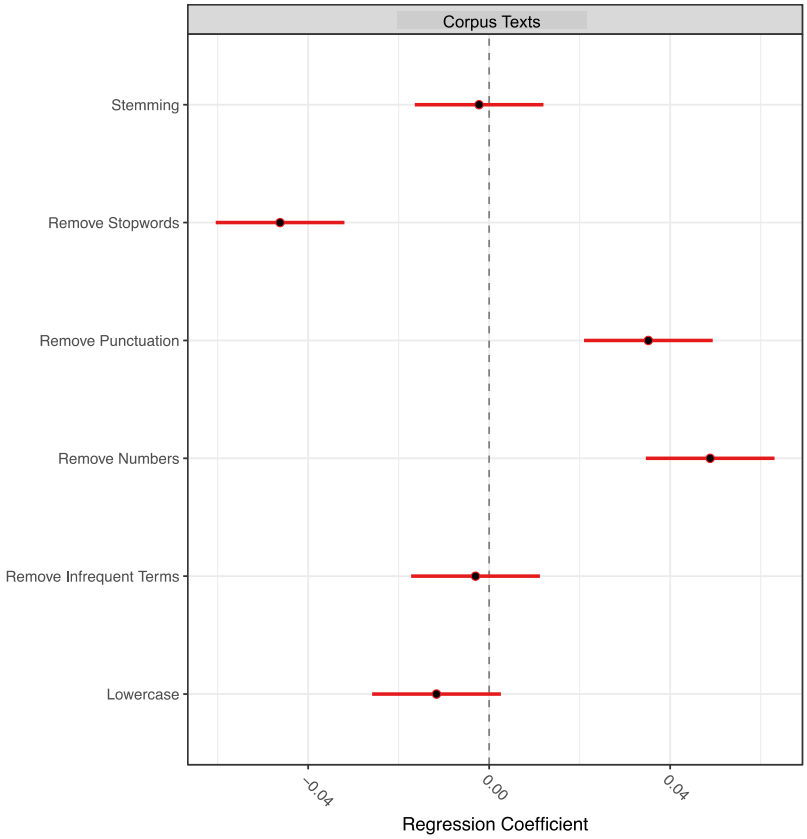

Figure S1. Regression Coefficients across Preprocessing Steps

Finally, we parsed each publication into pages, which served as our observational units, or documents, for the structural topic modeling (Grimmer and Stewart 2013). In section A.3, we evaluate the use of pages as documents. At this stage, a document-term matrix was created (Blei, Ng, and Jordan 2003) using stm.

A.2. Topic models: Results and interpretation

We analyze the document-term matrix with structural topic models (STM), implemented via stm (Roberts et al. 2016). These models include document metadata as covariates: documents' authorship, or the radical group that published the text, the year of publication, and the original language of publication. Our main results are generated by modeling a 10 topic solution (k). The selection of $k = 10$ is explained below.

The 10 topics are shown in Figure S2 as a display of topic labels and the 10 words with the

highest FREX scores. FREX scores are based on terms' probability of appearing under a topic and their exclusivity to that topic (Roberts et al. 2014). From these 10 topics, we select a subset of seven topics that are evocative of sociopolitical discourse (see the Varieties of Radical Rhetoric section in the main text): topics 1, 3, 4, 5, 6, 7, and 8.

1.	**War**	mujahideen, kill, attack, soldier, captur, enemi, oper, command, tank, area
2.	**Topic 2**	pepe, pace, hair, crush, bottom, fashion, mesa, dam, code, matt
3.	**Theology of jihad (struggle)**	god, jihad, quot, messeng, allah, brother, prophet, sheikh, bless, martyr
4.	**Local education & markets**	educ, thousand, provinc, news, train, respect, ministry, school, abdul, mullah
5.	**Afghan jihad (battle)**	mujahedeen, jihad, soviet, najib, kabul, regim, afghan, elect, presid, pakistan
6.	**Religious code**	movement, holy, life, great, human, scholar, women, quran, answer, world
7.	**International politics**	afghanistan, russian, countri, govern, unit, intern, russia, support, state, american
8.	**Social & political revolution**	social, uniti, polit, communiti, revolut, parti, west, exist, natur, system
9.	**Topic 9**	grant, sunday, catalog, prison, friday, inevit, servic, pose, pro, innov
10.	**Topic 10**	profil, breath, select, patent, post, bbc, threat, bar, tuesday, cost

Figure S2. Topic Labels and Top FREX Words
Note: Words have been stemmed.

Finding a subset of topics to be non-relevant is not surprising. Radical groups' writing will not solely reference sociopolitical discourse (Rana 2008), so not all topics will be relevant to our study on radicalism. For example, topic 10 appears to capture writing about the interviews that many *mujahideen* gave to the British Broadcasting Corporation (BBC). In addition, words can be clustered into a topic simply because of their lack of fit in other topics. These "residual topics" are often difficult to interpret—this is what we see with topics 2 and 9, with words such as "pace," "hair," "code," and "sunday," "catalog," and "friday," respectively. For further discussion on using a subset of topics, see Nielsen (2017).

The seven selected topics are grouped into two varieties of radical rhetoric, *subversion* and *reversion*, as discussed in the main text's Varieties of Radical Rhetoric section. The relative prominence of each rhetorical variant per document is then measured with a *rhetoric ratio score*. This score is a ratio of the prevalence of each rhetorical type's underlying topics. That is, in each

document, the prevalence of the topics in the set that defines radicalism of subversion is compared to the prevalence of the topics in the set that defines radicalism of reversion. In our case, the greater the prevalence of topics in the subversion group (i.e., topics 1, 3, 5, 7, and 8), the closer a document's rhetoric ratio is to one. The more a document is made up of topics in the reversion group (i.e., topics 4 and 6), the more a score approaches zero. Finally, each radical group is given a rhetorical ratio score by calculating the mean score of their documents.

A.3. Topic models: Validation

Topic models produce a measurement of a corpus. The quality of topic models' output should therefore be evaluated like other measures: by its reliability and validity (Quinn et al. 2010:216). Because any given topic model can be repeated (Nelson 2017),[4] discussion about assessing topic model results has focused on validity. In this section, we first present validation of our topic model results and then we offer a validation of the aggregation of topics into two rhetorical types.

A.3.1. *Validation of topics*

The validation of topic models typically entails assessing two related dimensions: whether the output is a "good" measurement of the corpus, which includes the selection of the topic solution (*k*), and whether it is credible, or the degree to which the topics "make sense" (Wilkerson and Casas 2017:533). We discuss each in turn.

There is no right answer to the number of topics that should be modeled (Roberts et al. 2014). As Grimmer and Stewart (2013:270) note, "the complexity of language implies that all methods necessarily fail to provide an accurate account of the data-generating process used to process texts. Automated content analysis methods use insightful, but wrong, models of texts to help researchers make inferences from their data." This is akin to there being no correct way to measure, say, "poverty." In studies examining poverty, selecting a measure of poverty follows from the theoretical insight that can be gained by the measure, as well as that insight's relevance to the research question.

With this property of topic models in mind, evaluating the selection of how many topics to model should be based on whether it provides "new and useful categorization schemes for text" (Grimmer and Stewart 2013:270). Or, as Quinn and colleagues (2010) put it, does the solution, which offers a view of the data at a particular granularity (Roberts et al. 2014:6), "work"? By this criterion, our selection of a topic solution (*k*) was good because it helped us develop a new understanding of radicalism.

To provide further confidence in the selection of a topic solution, the common practice has been to use expert knowledge to assess the output of different specifications of *k* (see Grimmer and Stewart 2013). For example, Quinn and colleagues (2010) and Lucas and colleagues (2015) recommend that researchers fit models with small, medium, and large values of *k*, and then "qualitatively" evaluate the usefulness of the varying results.

We followed this advice by making use of the groups' ratio scores. As discussed in the main text and in Appendix A, we consider all the groups that produced the texts in our corpus as radical. However, the scholarship on Afghanistan has long noted differences in the political objectives of these groups (e.g., Christia 2012; Roy 1986; Rubin 2002). For example, some supported the reinstatement of Afghanistan's exiled king, whereas others advocated for the creation of an Islamist state. Thus, the rhetoric ratio score should differ somewhat between some groups. For instance, "traditionalist" radicals should not have the same score as "Islamist" radicals. We

observed such logical differences in the relative scores based on the $k = 10$ model specification. Using k values of 7 and 13 produced results not in alignment with our knowledge, nor scholarship, of the case: the resulting group rhetorical ratio scores indicated the groups were largely indistinguishable from one another, which clearly contradicted the historical record. Consequently, we settled on an STM using a 10-topic solution.

Furthermore, the granularity of the topics estimated via a 10-topic solution was appropriate for conceptualizing thematic patterns and types of rhetoric within the discourse of radicalism, a central goal of our study. Fewer topics (i.e., $k = 7$) resulted in too broad topics—distinct concepts were "squished" together—whereas a solution greater than 10 (i.e., $k = 13$) produced topics that would have been more meaningful if combined with one another—a too fine-grained measurement of the corpus. In other words, the 10-topic solution provided a view of the corpus that helped us achieve our research goal, a key criterion for evaluating a measure's application (Grimmer and Stewart 2013; Nelson 2017; Quinn et al. 2010; Roberts et al. 2014).

Recent studies have supplemented the qualitative validation approach with a data-driven selection of k. One technique, proposed by Roberts and colleagues (2014) and recently implemented by Light and Odden (2017), compares the exclusivity and semantic coherence of topics for different solutions. Exclusivity refers to the extent to which frequent words are contained in a single topic. Coherence is the tendency for topics' most-probable words to co-occur together (Roberts, Stewart, and Tingley 2017). Topic solutions that generate higher exclusivity and coherence scores are understood to produce topics that are more semantically useful—they are consistent while also different from one another (Roberts et al. 2014).

Figure S3 shows the relationship between exclusivity and coherence for various candidate models of our corpus. We see that the 10-topic solution produces some of the most semantically useful results. A five-topic solution generates more coherent but less exclusive topics. Moving from five to 10 topics results in a large increase in exclusivity with a relatively small decrease of coherence. In contrast, moving from 10 to 15 topics results in a large reduction in coherence with relatively little increase in exclusivity. Continuing to increase the value of k incrementally increases the exclusivity, but, again, by weakening coherence. In practice, this means high-frequency words would be spread across ever more topics—words that otherwise might better belong in the same topic.

Our validation procedures give us confidence in our topic solution. A qualitative assessment based on the historical record of Afghanistan indicates that a 10-topic solution is best, as does the ultimate usefulness of this measurement. Moreover, an evaluation of candidate solutions' exclusivity and coherence—an evaluation conducted *after* we qualitatively selected our solution—also suggests the 10-topic solution efficiently generates one of the most semantically useful views of the corpus.

A second key issue when evaluating the quality of measurement is the selection of units, or documents. How might parsing the text into, say, paragraph- or page-length documents affect our inferences? The rule of thumb is to select a text length in which there is meaningful discontinuity (Grimmer and Stewart 2013). In our corpus, however, doing so is not straightforward. Our dataset contains a variety of kinds of text, throughout which there is no consistent use of break point. For example, some printed speeches run for pages without paragraph breaks. In light of this, we used page breaks to demarcate our documents, reasoning that, first, the length of a page roughly captured the ideas the authors opted to place together and, second, every publication in our corpus had page breaks.

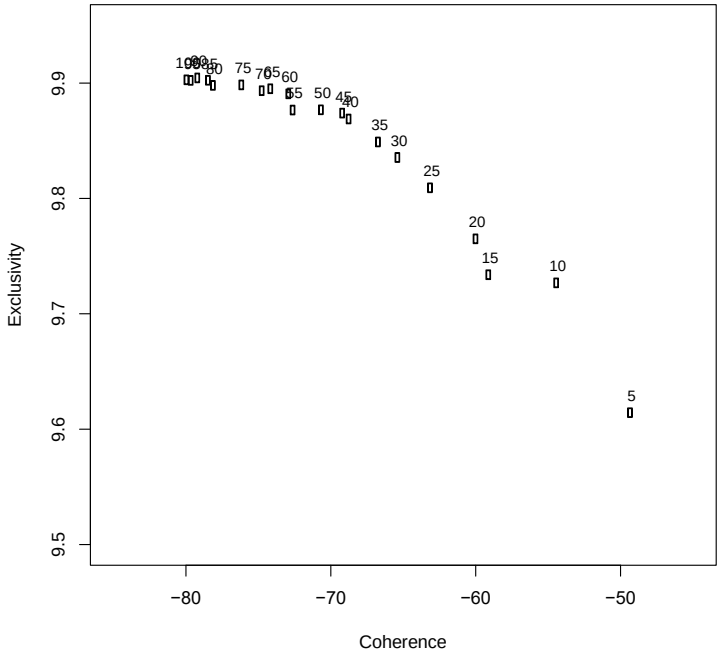

Figure S3. Exclusivity and Semantic Coherence of Topics for Different Topic Solutions
Note: Each point shows the exclusivity and coherence for a given topic solution, or number of topics. The topic solutions are indicated by the numbers assigned to each point.

To assess how the use of page-demarcated units might influence our findings, we first compare our main results to the results from a topic model drawing on documents of 500-word segments—a length of text that is about a page long, but not marked by page layouts. Table S1 presents the top 10 FREX words generated by these two models, with the main results listed in the second column. We observe consistency in the topics. In addition, we find largely similar FREX words rankings, although the extent of change in the rankings varies across topics.

Having established that models using documents defined as a page or 500 words produce similar results, we turn to assessing the output produced from documents of other lengths. Because, as mentioned earlier, not all publications in our corpus contain paragraph breaks, we create documents that are 150 and 350 words long and use these in the model. Then we compare the exclusivity and semantic coherence of these models' results to those of existing results. Figure S4 shows that modeling a 10-topic solution with documents 500 words long, or about a page length, maximizes exclusivity and coherence. Comparing the results shown in Figure S4 to those in Figure S3 indicates that using page-demarcated documents produces topics that are more semantically coherent but slightly less exclusive than topics generated using 500- and 350-word long documents.

Table S1. Comparison of Top FREX Words

Topics	Page-length documents	500-word documents
War	mujahideen, kill, attack, soldier, capture, enemy, command, tank, area	mujahideen, kill, soldier, attack, capture, enemy, tank, operation, destroy, wound
Theology of jihad (struggle)	god, jihad, quote, messenger, allah, brother, sheikh, bless, martyr	allah, god, prophet, messenger, brother, jihad, almighty, bless, heart, martyr
Local education & markets	education, thousand, province, news, train, respect, ministry, school, abdul, mullah	education, train, children, studies, school, women, university, student, institution, province
Afghan jihad (battle)	mujahideen, jihad, soviet, najid, regime, afghan, election, president, pakistan	mujahideen, soviet, jihad, afghan, regime, kabul, refuge, najib, peshawar, volunteer
Religious code	movement, holy, life, great, human, scholar, woman, quran, answer, world	human, social, societies, communities, right, system, principle, mean, nature, life
International politics	afghanistan, russian, countries, govern, unit, international, russia, support, state, american	russian, movement, war, russia, muslim, defeat, struggle, face, armies, victory
Social & political revolution	social, unity, political, communities, revolution, parties, west, exit, nature, system	parties, council, elect, leader, iran, unit, leadership, solution, member, decision

Note: Table shows the top 10 FREX words generated from two models using different document definitions. The second column presents the main results: topics and words generated from a model drawing on page-demarcated documents. The third column presents the results of a model drawing on documents 500 words in length. Endings have been added to stemmed terms based on the usage of words in highly associated documents.

Finally, we evaluate the credibility of our topic model results. To do so, we conduct a test of predictive validity (Quinn et al. 2010). Results are high in predictive validity if they vary along features external to the modeling process in expected ways. For example, we gain confidence in the credibility of topics found in U.S. senators' communication with constituents if the prevalence of specific topics in a senator's communications corresponds to that senator's committee assignment (Grimmer and Stewart 2013).

Our test compares the estimated prevalence of topics across radical groups. If our results are credible, the proportion of certain topics should differ based on the groups' activities. For example, a topic that captures battlefield violence should be more prevalent in the writing by a militant group routinely engaging in armed fighting.

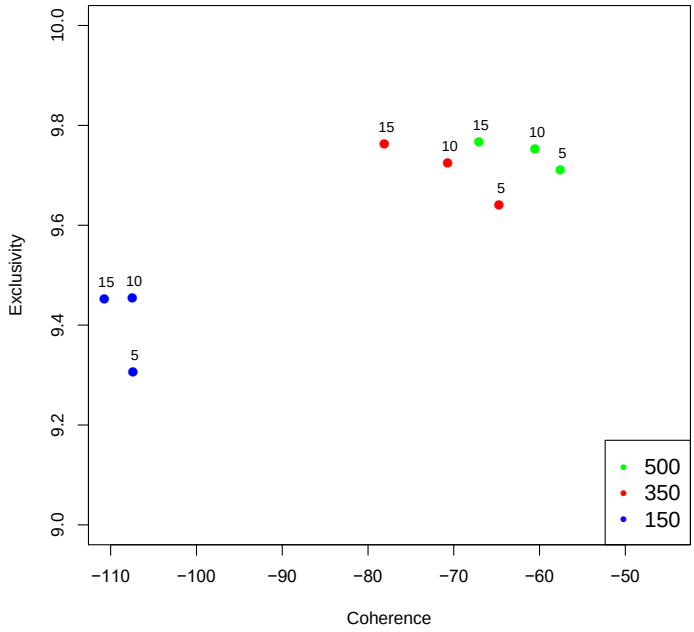

Figure S4. Exclusivity and Semantic Coherence of Topics for Different Document Lengths and Topic Solutions
Note: Each point shows the exclusivity and coherence for a given topic solution, or number of topics. The topic solutions are indicated by the numbers assigned to each point. Colors indicate the relationship for topic solutions when using documents of 500, 350, and 150 words long.

We test three topics across three groups, the Cultural Council of Afghanistan Resistance (CCAR), *Jam'iyyat-i Islami*, and *Hizb-i Islami* (Gulbuddin). We selected these groups because our corpus contains similar amounts of their publications across time. The first topic, "war," which comprises language about battlefield violence (e.g., "kill," "attack," "soldier," "tank"), should be most prevalent in the documents produced by *Jam'iyyat*, the group that was most active in the fighting (Rubin 2002). The second topic, "theology of jihad," made up of high-frequency words like "god," "message," and "prophet," should be more common in the writings of *Hizb-i Islami*. Among our selected groups, *Hizb-i Islami* most prioritized its Islamist agenda. As Roy (1986:133) observed at the time, its "power base did not extend beyond the network of Islamists" and "it regarded the Islamic revolution as being more important than the war" against the communists. Finally, the third topic, "local education & markets," should be most prevalent in the CCAR's publications. The CCAR was not militarized; it focused on political and social issues.

Figure S5 presents the results of our test. Our topic model output varies in the ways we would expect. The "war" topic is more often estimated to have a greater prevalence in the writings of *Jam'iyyat*; the estimated proportion of "theology of jihad" is most frequently greater in the documents produced by *Hizb-i Islami*; and the "local education & markets" estimate is most often greatest in CCAR's publications. These findings suggest our STM results are credible. In summary, the assessments of measurement quality and credibility together give us confidence in the validity of our topic model results.

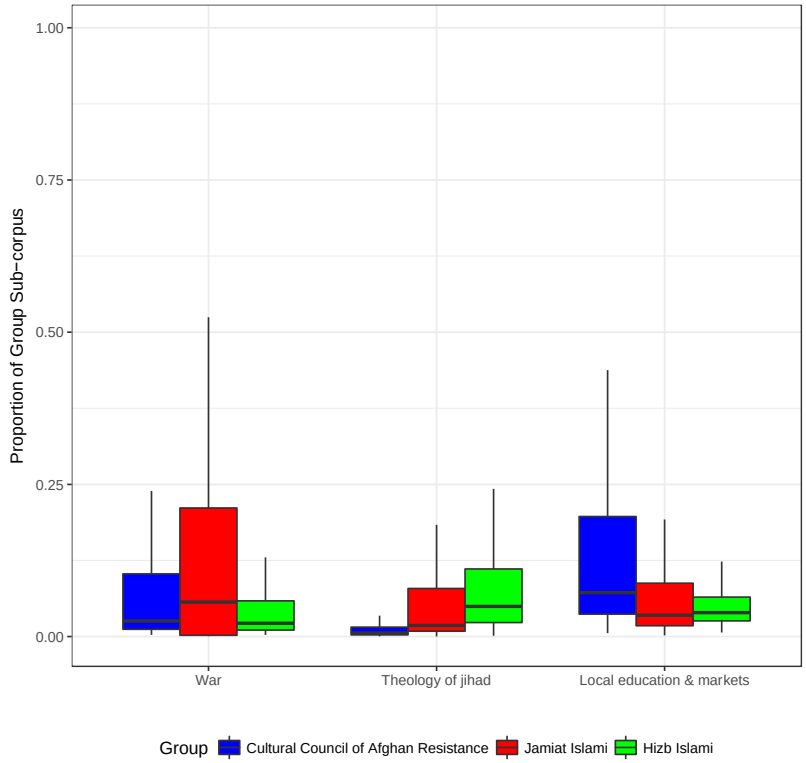

Figure S5. Prevalence of Selected Topics by Group
Note: The estimated topic proportion of specific topics vary across groups in expected ways. The boxes denote the median and 25 to 75 percent quartile range. Points are outliers.

A.3.2 *Validation of rhetorical variants*

We also validate the aggregation of our topics into two rhetorical varieties using the criteria of credibility (i.e., does it "make sense"?) and quality of measurement (i.e., does it capture an important aspect of the data well?) (Wilkerson and Casas 2017:533). The assessment of the latter is provided by the theoretically driven computational abductive analysis, as explained in the main text. This section is used to evaluate the former.

As in the preceding section, we check the aggregation's credibility using predictive reasoning: the results of the aggregation should correspond to external events, categories, or findings in expected ways (DiMaggio, Nag, and Blei 2013; Grimmer and Stewart 2013). Because one hallmark of our rhetorical types is that they differ in how they portray radicals' *connectivity* to the future, or the logic by which radicals move toward their imagined future (see the Varieties of Radical Rhetoric section in the main text), we compare their variation in language to the variation observed in other depictions of connectivity.

In her study of language used by environmental policymakers and activists, Mische (2014) identified sets of words associated with kinds of connectivity. One set, "advance," "against," "attack," "fight," and "must," indicates a call to action—an engagement in a struggle to reach a certain future. This kind of connectivity is akin to our radicalism of subversion. A second set of words suggests that a future will happen under certain conditions, similar to how a radicalism of

reversion encourages listeners to turn inward toward themselves and their families and communities to engender a desired future. These words are "can," "might," "likely," "might," "will," and "would." We use a selection from the first and second sets of words to denote the rhetorics of subversion and reversion, respectively.

Our validation test entails two kinds of comparisons. First, we compare the frequency of the exogenously defined subversion terms in the subversion text to their frequency in the reversion text. At the same time, we compare the frequency of the reversion terms in the subversion text to their frequency in the reversion text. If we have developed a good measurement of our theorized rhetorics, the subversion terms should be more frequent in the subversion text, and the reversion terms should be more frequent in the reversion text. In addition, we conduct these word frequency comparisons with random aggregations of topics, or what can be thought of as placebo rhetorics.

Second, we compare (1) how well the word frequency comparisons with our theoretically justified aggregations align with expectations to (2) how well the word frequency comparisons done with placebo rhetorics align to expectations. The comparisons done with the theoretically justified aggregations—the rhetorics of subversion and reversion—should fit expectations better than the comparisons using the placebo aggregations.

To conduct our test, we first identify the documents that have an estimated topic proportion greater than 75 percent for each topic. These documents are then used to create a single corpus, which is subsequently preprocessed as described in Section A.1 and subset to only contain common English words. The overall results remain the same if the word frequencies are normalized using the entire corpus. We then combine the documents highly associated with topics 1, 3, 5, 7, and 8 into a subversion text. The documents highly associated with topics 4 and 6 are compiled into a reversion text.

The placebo texts are formed by compiling all possible random combinations of topics. For example, the documents highly associated with topics 1, 5, and 6 might be combined into a "rhetoric," captured by a "T1-T5-T6" text. The rhetorics reflected in the placebo texts might be arbitrarily labeled as "subversion" for ease of interpretation, but they could have been labeled as examples of reversion. We present the results of example placebos in Table S2, but we also calculate the mean result for all placebos.

Table S2 shows the results of our comparisons. The word frequencies are percentages of the total words in each aggregation's compiled text. For example, a value of .01 for the term "advance" in the subversion compilation of text indicates that .01 of all words in the compilation are "advance." The frequencies of our selected terms differ across the theoretically justified subversion and reversion rhetorics in the expected ways. All the subversion terms are more common in the subversion rhetoric than in the reversion rhetoric. Similarly, all the reversion terms appear more frequently in the reversion text than in the subversion text. Some terms appear twice as often in the appropriate rhetoric.

Table S2. Percent of Test Terms in Each Rhetorical Category

		Rhetorics		Random selection 1		Random selection 2	
		Subversion Topics 1, 3, 5, 7, 8	Reversion Topics 4, 6	"Subversion" Topics 1, 5, 6	"Reversion" Topics 3, 4, 7, 8	"Subversion Topics 3, 6	"Reversion" Topics 1, 4, 5, 7, 8
Subversion test	advance	**0.01**	**0.01**	0.00	0.01	0.00	0.01
	against	**0.19**	**0.16**	**0.36**	**0.05**	0.07	0.31
	attack	**0.40**	**0.01**	**0.76**	**0.01**	0.01	0.69
	fight	**0.05**	**0.02**	0.04	0.06	**0.07**	**0.03**
	must	**0.07**	**0.05**	0.03	0.09	**0.10**	**0.03**
Reversion test	can	**0.10**	**0.37**	0.14	0.13	0.18	0.10
	likely	**0.00**	**0.01**	**0.00**	**0.01**	0.01	0.00
	might	**0.01**	**0.03**	0.02	0.01	0.02	0.01
	will	**0.32**	**1.01**	0.48	0.38	0.43	0.42
	would	**0.13**	**0.22**	0.22	0.08	**0.08**	**0.20**

Note: The theoretically justified rhetorical categories (i.e., aggregations of topics), rhetorics of subversion and reversion, have a greater number of comparative word frequencies that differ as expected than the aggregations of randomly selected topics. The values in the cells indicate the percent of all words in the rhetorical categories that are the test terms. Values have been rounded. The bold pairs of percentages across each pair of theorized or placebo rhetorics indicate a difference in the expected direction, although the placebo results should also be evaluated with labels switched. See Section A.3.2 for details.

In the second step of our test, we find that the theoretically justified aggregations differ in expected ways more often than the random aggregations. The first example placebo comparison (i.e., topics 1, 5, and 6 versus topics 3, 4, 7, and 8; "Random selection 1" in Table S2) achieves three out of ten comparisons correctly, or, when the arbitrarily assigned labels are switched, seven out of ten comparisons align with expectations. In the second example placebo comparison (i.e., topics 3 and 6 versus topics 1, 4, 5, 7, and 8; "Random selection 2" in Table S2) three out of ten expectations align with expectations, or, when the labels are switched, seven out of ten match expectations. Thus, two of the example placebos come close (70 percent success) and the other two perform poorly (30 percent success) while the rhetorics of subversion and reversion achieve 100 percent success in this test. The mean success rate of all possible aggregations is six out of 10 (60 percent) with a standard deviation of 1.41. The success rate of the theoretically justified categories is greater than one standard deviation of the placebos' mean success rate. These results suggest the rhetorics of subversion and reversion better align with previously observed rhetorical treatments of connectivity (Mische 2014) than do other possible aggregations of our topics.

In addition, we check how using individual topics instead of aggregated rhetorics would affect our insights regarding the adoption of rhetoric (i.e., the second part of our study). Overall, we arrive at the same conclusion. These results are presented in Section D.1.

A.4 Topic models: Robustness

An important consideration when interpreting STM output is how robust the topics are to the omission of covariates. When producing our initial STM results, we did not consider how the type of publication (e.g., magazine, newspaper, speech) may influence the topical content. To check for this potential effect, we re-estimated our model with the addition of publication type to the original set of covariates. We found no major differences in the output of the original and

updated model. The rank ordering of FREX words in the updated model was slightly different (Figure S6), but the content of the topics remained nearly the same (compare to Figure 1 in the main text). In light of these findings, we used the original output for the main analysis.

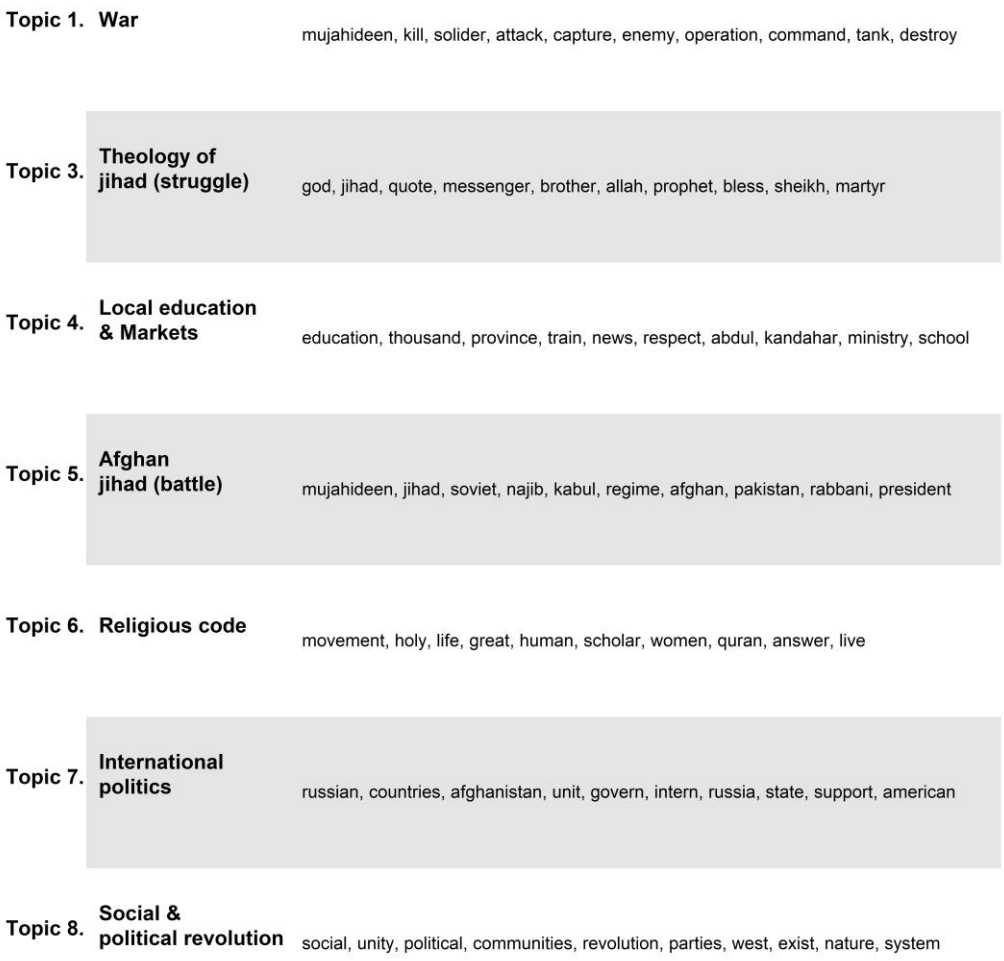

Figure S6. Labels and Top FREX Words for Selected Topics Estimated from STM with Publication Covariate

Note: Figure shows the content of topics estimated with an STM using the original covariates and the addition of a covariate for publication type. The content is nearly identical to the original output (Figure 1 in the main text). Endings have been added to stemmed terms based on the usage of words in each topic's highly associated documents.

Part B. METHODOLOGICAL DETAILS OF THE FRACTIONAL LOGISTIC ANALYSIS

The initial fractional logistic models utilize two main sources of data. One, our dependent variable, is a document-level ratio of subversion to reversion rhetorics. The ratio is determined for each document in our corpus by comparing the proportion of a document's text associated with topics underlying each rhetorical variant (see Section A.2). In our case, the higher the value of the ratio, the more a document's discourse reflects a radicalism of subversion. To gain further confidence in our conclusions, we subsequently redo the analysis using the prevalence of individual topics, rather than aggregated rhetorics, as the dependent variable. These results are presented in Part D.

The second source of data, the Uppsala Conflict Data Program (UCDP),[5] informs our independent variables. Specifically, we use the UCDP datasets containing information about external support (Högbladh, Pettersson, and Themnér 2011) and groups' casualties and number of fighters (Sundberg, Eck, and Kreutz 2012), as well as the affiliated ACD2EPR dataset[6] (Vogt et al. 2015; Wucherpfennig et al. 2012) for ethnic and sectarian affiliation.

Observations are annual-level because the UCDP data on our independent variable of interest—whether a group receives external financial support—is per annum. As a result, we average the rhetoric score for each group's documents in a given year. Our explanatory variable of interest, whether or not a group received external financial support, is binary.

We conduct four specifications of a fractional logistic model using Stata 14.0. We first estimate a bivariate model of external support's effect on the rhetoric ratio, then a model that includes the covariates of group casualties and number of fighters. The former is measured in the hundreds of persons, the latter in the thousands of persons. Recall that the STM analysis accounted for the effects of group authorship, year, and language. Yet, as a robustness check, we include a specification using group fixed effects, as well as one including both group and temporal fixed effects (see Part D). For the temporal fixed-effect specification, we divide the years between 1979 and 2001 into five periods established in the literature on the Afghanistan conflict (Christia 2012; Rubin 2002): the early Jihad, the late Jihad, the intra-*mujahideen* civil war, the early Taliban period, and the late Taliban period. The results are reported as mean marginal effects, derived using the margins dydx () command. These coefficients should be interpreted with no funding (i.e., funding = 0) as the base level. Standard errors are clustered by group and period.

As a further robustness check, we drop one prominent radical group's observations and re-estimate the fractional logistic models with group and period fixed effects. These results are reported in Part D.

Next, we estimate a second set of fractional logistic models using a multiplicative interaction term. Data on the moderator, global oil price, is obtained from the International Monetary Fund's public database. The IMF reports oil price as a simple average of three price measures—Dated Brent, West Texas Intermediate, and the Dubai Fateh—in U.S. dollars, indexed to 2005. We interact this moderator with the origin of groups' external support. External support is labeled as being from one of two places: (1) Saudi Arabia or other predominately Arab states in the Persian Gulf region, oil-rich countries or "petrostates," and (2) any other country. We identify groups' patron with data from the UCDP External Support dataset (Högbladh et al. 2011) and by the assumption that groups with large numbers of Arab fighters receive support from Arab Persian Gulf states. This assumption is used to identify four groups as receiving support from the Arab Persian Gulf: the Haqqani Organization, Miramshah College, *Jami'ah al Da'wah ila al-Quir'an wa al-Sunnah,* and *Maktab al-Khidamat*.

Iran, an oil-producing country, also provided external support to some groups in Afghanistan, but we only select groups with ties to the Arab Persian Gulf into the "treated," or oil-state supported, category because monetary resources in Iran are relatively less sensitive to global oil prices due to sanctions imposed on its petroleum industry and international trade after 1979 (and expanded in 1995). After the imposition of sanctions, Iran's oil production collapsed from pre-1979 levels, making oil production a smaller part of its economy. In addition, the sanctions meant Iran's oil production and revenue did not respond to oil price as production and revenue did in the Arab Persian Gulf. For example, during the rise in West Texas Intermediate prices in the early 1980s and the (expected) corresponding increase in production by oil-rich Arab states, Iran's production decreased.

Because oil price is recorded at monthly intervals, we use month-level observations in the multiplicative interaction portion of the analysis. This means we return to our original rhetoric ratio values—each group is assigned a mean rhetoric score at monthly intervals. Doing so gives us more observations and greater statistical power. The disadvantage is that we cannot include the annual measures of group casualties and number of fighters as covariates. However, we use fixed effects, which our initial model results indicate produce coefficients in the same direction and at the same level of significance as the specifications with the covariates (see Table 3 in the main text and Table S4 in Part D).

For the multiplicative interaction specifications, we use the interflex[7] package for R (Hainmueller, Mummolo, and Xu 2019). The model uses a simple binning estimator. This approach offers several advantages, as reported in Section 4.1 of Hainmueller and colleagues (2019). We highlight two: first, it serves as a formal test of the validity of the linear interaction effect assumption imposed by the standard model (the assumption is validated in our case) and, second, conditional marginal effects are estimated at typical values of the moderator (in our case, the medians of the first, second, and third terciles of the moderator).The dependent variable, the monthly rhetoric ratio, is a continuous value falling between zero and one. The focal independent variable is dichotomous, indicating whether a group received support from an Arab state in the Persian Gulf in a given month. The moderator is the price of oil, measured in U.S. dollars. This last unit of measure is appropriate because marginal effects do not provide a very good approximation of the effect of a one unit increase in the independent variable when the variable's unit is large, but they *do* provide accurate approximations when the units are small, such as a dollar.[8] Standard errors are clustered by group and period. We report the findings as mean marginal effects.

To evaluate the impact of our assumption that groups with many members from the Arab Persian Gulf states receive support from those states, we estimate the fractional logistic multiplicative interaction model when the category of groups supported by the Arab Persian Gulf states only includes, first, groups that explicitly received support from Saudi Arabia (as recorded by UCDP) and, second, only the four groups we assume to have ties to the Arab Persian Gulf based on their high numbers of Arab fighters (i.e., the Haqqani Organization, Miramshah College, *Jami'ah al Da'wah ila al-Quir'an wa al-Sunnah*, and *Maktab al-Khidamat*). This check verifies our identification of support from the Arab Persian Gulf (see Part D).

Finally, we check the impact of our assumption that radical backers in Saudi Arabia and the Gulf base their decision to offer support on the wealth they expect to receive from current oil prices. Under this assumption, we do not lag the price of oil. However, because we do not have strong evidence in support of this assumption, we also estimate our model using oil prices lagged by one month and six months. The overall finding remains the same (see Part D).

Part C. EFFECT OF SUPPORT FROM SAUDI ARABIA OR ARAB GULF STATES

Table S3 displays the coefficients from the fractional logistic multiplicative interaction models. The first model estimates the effect of support from Saudi Arabia or the Arab Persian Gulf, as well as that of the moderator, global oil price, without the interaction term. We find a significant positive effect for origin of support ($p < .05$).

The second model adds the interaction term; its coefficient is positive and significant ($p < .01$), as shown in Figure 3 in the main text. The main effect of oil price—or, the effect of oil price on groups that have support from outside SA or the Gulf—lessens the rhetoric of subversion. This might be because of the redistribution of wealth when oil price increases. Namely, residents and governments of non-oil producing states have to spend more money on oil when prices increase, resulting in less support available to radicals. This decreasing availability of (potential) external support produces a corresponding decrease in the rhetoric of subversion, as our argument suggests.

The remaining models serve as robustness checks, as explained in Part D; the results are discussed in D.3.

Table S3. Effect of Support from Saudi Arabia (SA) or Arab Gulf States on Radical Rhetoric of Subversion, Moderated by Oil Price

	(1)	(2)	(3)	(4)
Price of oil	0.000726	-0.00144**	0.000353	-0.00157**
	(0.00100)	(0.000213)	(0.00123)	(0.000245)
Support from SA or Gulf	0.0809*	-0.0377 (0.0581)		
	(0.0391)			
Oil Price X Support from SA or Gulf		0.00343**		
		(0.00116)		
Support from SA			0.103**	-0.0273
			(0.0397)	(0.0776)
Oil Price X Support from SA				0.00367*
				(0.00172)
Support from Gulf				
Oil Price X Support from Gulf				
Group FEs	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes
N	16,932	16,932	12,754	12,754
Pseudo R^2	0.046	0.047	0.022	0.022

Note: Coefficients are mean marginal effects derived from a fractional logit model. Standard errors, in parentheses, are clustered by group and period.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

Part D. ADOPTION OF RHETORIC: ROBUSTNESS CHECKS

D.1 Fractional logistic regression

Our first robustness check entails modeling the bivariate relationship between external support and the rhetoric of subversion using, first, group fixed effects (FE) and, second, group and temporal FE. We estimate this specification as a robustness check because the STM analysis accounts for the effects of group authorship, publication year, and language of the text. Table S4 presents mean marginal effects of the FE specifications, with no support as the base level.

When including group FE, receiving funding is associated with an increase of .232 ($p < .01$) in our rhetoric ratio. When using both group and temporal FE, the increase is .257 ($p < .01$). Considering that the mean rhetoric ratio is about .65, these increases are substantial. These results mirror the findings presented in the main text.

Table S4. Effect of External Support on Radical Rhetoric of Subversion, with Fixed Effects

	(1) Rhetoric Ratio	(2) Rhetoric Ratio
External support	0.232**	0.257**
	(0.0787)	(0.0904)
Constant	0.545**	0.726**
	(0.101)	(0.0908)
Group FEs	Yes	Yes
Period FEs	No	Yes
Observations	12,802	12,802
Pseudo R^2	0.019	0.021

Note: Coefficients are mean marginal effects derived from a fractional logit model. Standard errors, in parentheses, are clustered by group and period.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

Our second robustness check entails removing some prominent groups from the dataset. The findings are largely robust to these exclusions (Table S5). Excluding *Jam'iyyat*, *Jabha*, and *Hizb-i Islami* (Gulbuddin faction) does not change the direction and significance of the estimate, and only has a minor impact on the effect size. This check suggests the relationship between external support and rhetoric is not occurring within a specific group. The exception is the Taliban: when Taliban observations are removed, the effect becomes non-significant, although the direction of the effect remains positive. This change might be due to greater variance introduced by removing the relatively large number of observations (n = 11,181). However, it is worthwhile to consider that the Taliban might be an exemplary case of our argument—they did not receive as much external support as the *mujahideen* groups did during the Jihad (Zaeef 2010).

Table S5. Effect of Support on Rhetoric of Subversion, Excluding Prominent Groups

	(1) No Jam'iyyat	(2) No Jabha	(3) No Hizb-i Islami (G)	(4) No Taliban
External support	0.204**	0.257**	0.268**	0.245
	(0.0658)	(0.0906)	(0.0959)	(0.278)
Constant	0.839**	0.726**	0.715**	0.738**
	(0.0658)	(0.0910)	(0.0963)	(0.278)
Group FEs	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes
Observations	6,729	12,443	9,877	11,181
Pseudo R^2	0.029	0.020	0.027	0.005

Note: Each column presents the results of models estimated without observations from one key group. Coefficients are mean marginal effects derived from a fractional logit model. Standard errors, in parentheses, are clustered by group and period.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

Third, we seek to gain further confidence in our conclusion that radicals' networks of support affect their use of language by estimating the effect of external support on the prevalence of individual topics. Doing so also sheds light on the extent to which our insights depend on our (theoretically informed) aggregation of topics into rhetorics.

We estimate seemingly unrelated regressions (SUR), or multiple equations related to one another through a correlation in the errors. This way, we capture the fact that resources can have different but linked influence on the prevalence of each topic. In other words, we estimate the effect of resources on the prevalence of a given topic, relative to all other topics, while taking into account that the effect may be related to the effect of resources on each of the other topics.

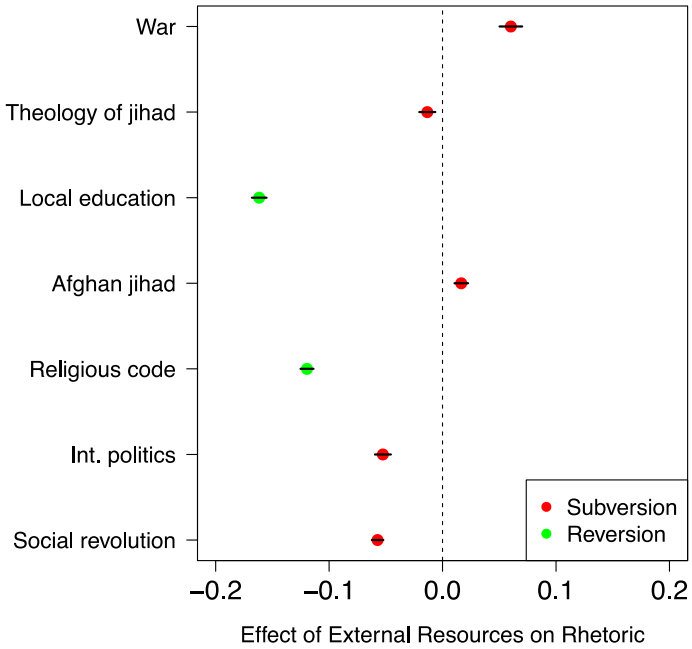

Figure S7. Effect of Support on the Prevalence of Topics
Note: Figure shows the estimated effect of external resources on the prevalence of topics using seemingly unrelated regressions (SUR). The color of the point estimates indicates whether the topic forms part of the subversion or reversion rhetoric. 95 percent confidence intervals are indicated by the black bars.

Figure S7 presents the results. We find that receiving external support increases the prevalence of war-like topics. At the same time, groups receiving external support decrease their use of topics forming the rhetoric of reversion. There is also a decrease in some of the subversion topics, but they decrease relatively less in comparison to the reversion topics. Overall, these findings align with the expectations of our argument and main conclusion: receiving external support helps shift radicals' rhetoric toward subversion and away from reversion. However, in light of the evidence of heterogeneous effects across topics, we recommend that future research further explores the relationships between resources and individual topics, as well as between topics and rhetorics.

D.2 Accounting for groups' geographic area of activity

Between 1979 and 2001, most, if not all, of Afghanistan's radical groups had representatives of their leadership abroad, in relatively easy to access urban centers, such as the Pakistani cities of Peshawar and Quetta. However, many members of the groups' mid-tier leaders ("commanders") and soldiers lived and operated within Afghanistan, which has large portions of rugged and inhospitable terrain that made travel, transportation, and communication challenging. Consequently, while a group's leadership may have had certain relationships of support, their colleagues in the field could have experienced differing access to resources.

To account for this possibility in our analysis, we re-conducted the moderation analysis with the inclusion of a covariate reflecting groups' geographic area of activity on a monthly basis. To create this variable, we drew on data on violent incidents involving groups that were contained in the UCDP dataset[9] (see Part B). If a group was active in an attack located in a region of Afghanistan in a given year (the temporal unit used by UCDP), the group was coded as operating in that region during each month of that year. We apply observations at the annual level to months because insurgent groups are often active in an area before and after a violent incident may be recorded by observers. In addition, fighting in Afghanistan followed (and still largely follows) a seasonal pattern—insurgents established areas of operations during the warmer months and regrouped during winter. We used the commonly known regions of Afghanistan: north, west, central, east, and south. Including the area of operation covariate constrains our analysis to 2,480 monthly observations of groups and their rhetoric. Our main analysis relied on 23,412 monthly observations of groups and their rhetoric.

Our main results are robust to including data on groups' spatiotemporal location. Figure S8 shows that the effect of support from SA or the Arab Gulf states, moderated by oil price, continues to be positive, and still increases as oil prices rise. The confidence intervals of our estimates are much larger than in our main analysis because of the substantial reduction in observations. Figure S9 shows the results when we only use observations that are missing spatiotemporal data (n = 17,324). We obtain findings similar to both the main analysis and the analysis including the spatiotemporal covariate. The two new models presented in this section include group and period fixed effects (see Section D.1).

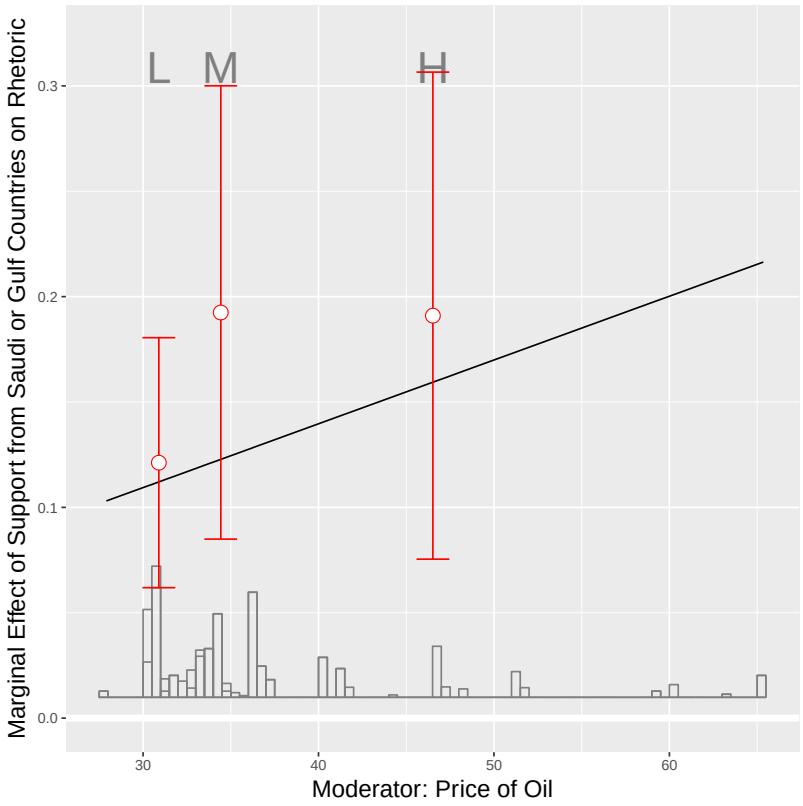

Figure S8. Effect of Gulf Support on Rhetoric, Moderated by Oil Price and Including Group Area of Activity

Note: The dark line and gray 95 percent confidence interval band depicts the conditional marginal effect of support from SA or Arab Gulf states across oil price levels estimated by the standard multiplicative interaction model. The point estimates with 95 percent confidence interval bars represent the conditional marginal effects from a binning estimator at the medians of the three moderator-value bins, the first tercile (L), the second tercile (M), and the third tercile (H). The stacked histogram shows the distribution of the moderator (oil price in U.S. dollars; 2005 = $100), with the red and white shaded bars indicating the distribution of the moderator in the treatment and control groups, respectively.

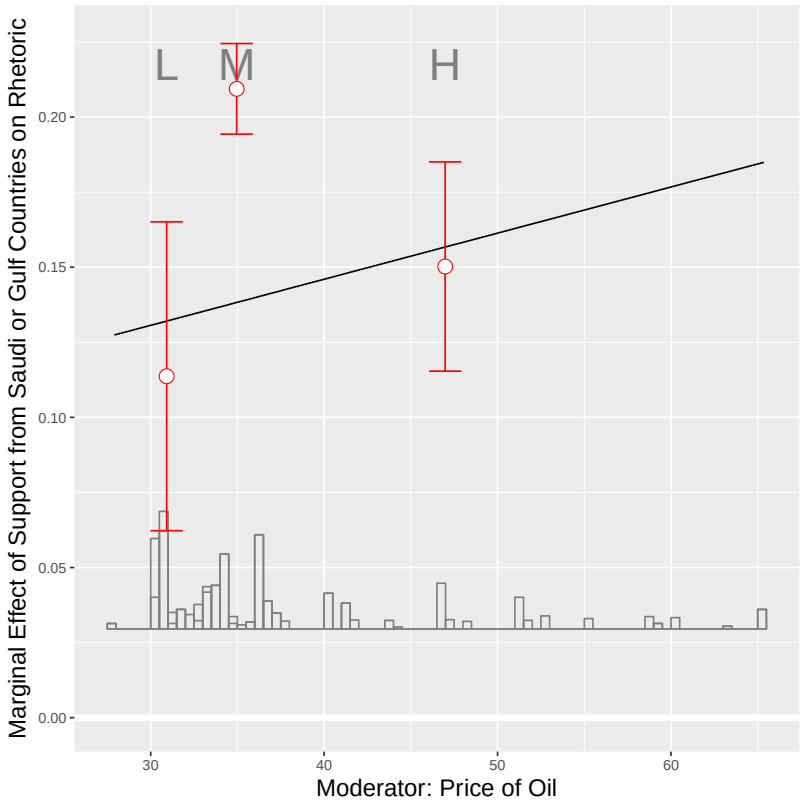

Figure S9. Effect of Gulf Support on Rhetoric, Moderated by Oil Price, Using Observations Missing Area of Activity
Note: The dark line and gray 95 percent confidence interval band depicts the conditional marginal effect of support from SA or Arab Gulf states across oil price levels estimated by the standard multiplicative interaction model. The point estimates with 95 percent confidence interval bars represent the conditional marginal effects from a binning estimator at the medians of the three moderator-value bins, the first tercile (L), the second tercile (M), and the third tercile (H). The stacked histogram shows the distribution of the moderator (oil price in U.S. dollars; 2005 = $100), with the red and white shaded bars indicating the distribution of the moderator in the treatment and control groups, respectively.

D.3 Support from oil-rich Arab states

The third through sixth models presented in Table S3 serve as robustness checks of the main analysis of how support from oil-rich Arab states affects rhetoric, as explained in Part B. These models show that interacting origin of support and oil price results in positive significant coefficients, whether considering support only from Saudi Arabia (see also Figure S10) or support only from other Arab Gulf countries. The latter observations are based on the assumption that a high rate of Arab fighters in a group is equivalent to that group receiving support from Saudi Arabia or Arab Gulf states (see the Analytic Strategy section in the main text). Figures S11 and S12 show that we achieve results similar to that of our main model (Model 2, Table S3) when oil price is lagged by one or six months.

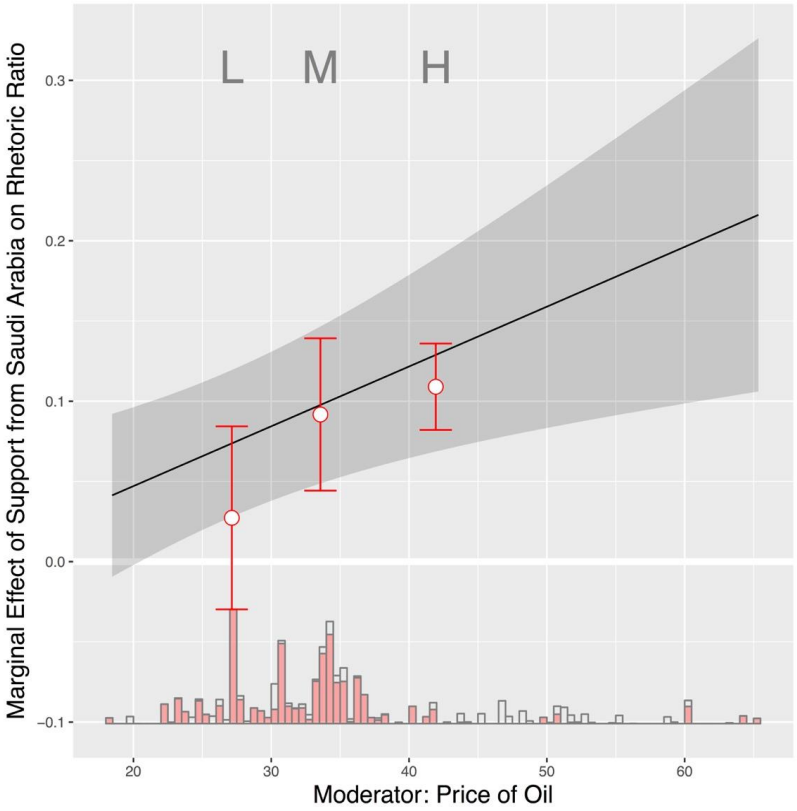

Figure S10. Effect of Support from Saudi Arabia on Rhetoric of Subversion, Moderated by Oil Price
Note: The dark line and gray 95 percent confidence interval band depicts the conditional marginal effect of support from SA or Arab Gulf states across oil price levels estimated by the standard multiplicative interaction model. The point estimates with 95 percent confidence interval bars represent the conditional marginal effects from a binning estimator at the medians of the three moderator-value bins, the first tercile (L), the second tercile (M), and the third tercile (H). The stacked histogram shows the distribution of the moderator (oil price in U.S. dollars; 2005 = $100), with the red and white shaded bars indicating the distribution of the moderator in the treatment and control groups, respectively.

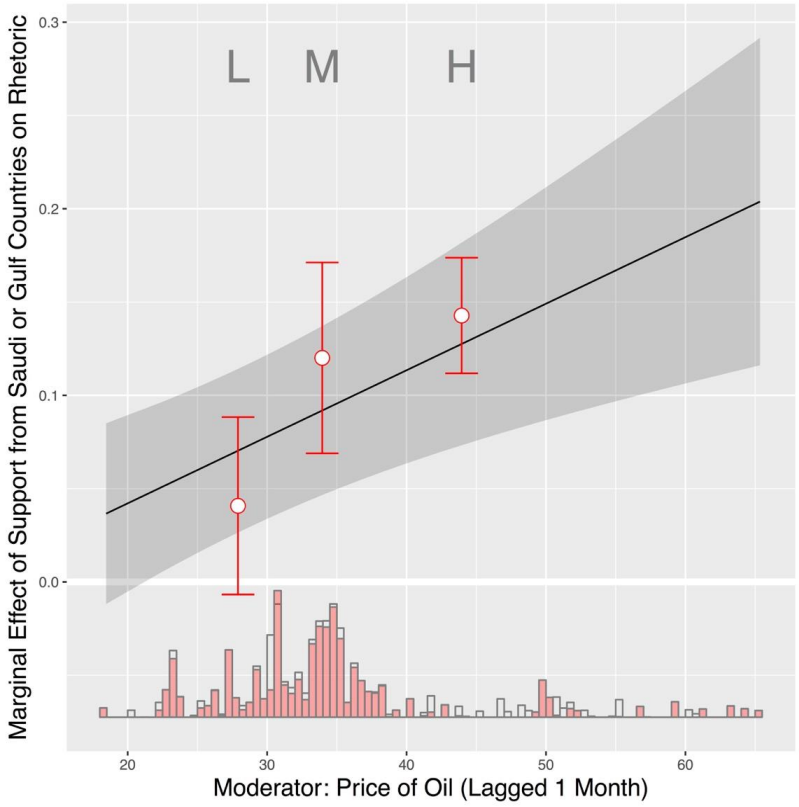

Figure S11. Effect of Support on Rhetoric of Subversion, Moderated by Oil Price Lagged One Month
Note: The dark line and gray 95 percent confidence interval band depicts the conditional marginal effect of support from SA or Arab Gulf states across oil price levels estimated by the standard multiplicative interaction model. The point estimates with 95 percent confidence interval bars represent the conditional marginal effects from a binning estimator at the medians of the three moderator-value bins, the first tercile (L), the second tercile (M), and the third tercile (H). The stacked histogram shows the distribution of the moderator (oil price in U.S. dollars; 2005 = $100), with the red and white shaded bars indicating the distribution of the moderator in the treatment and control groups, respectively.

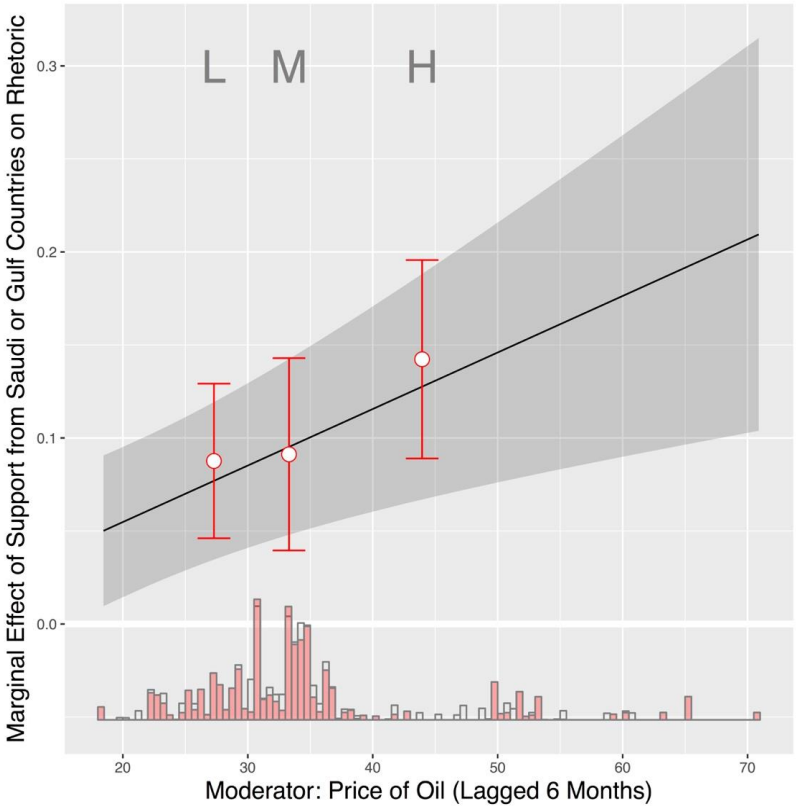

Figure S12. Effect of Support on Rhetoric of Subversion, Moderated by Oil Price Lagged Six Months
Note: The dark line and gray 95 percent confidence interval band depicts the conditional marginal effect of support from SA or Arab Gulf states across oil price levels estimated by the standard multiplicative interaction model. The point estimates with 95 percent confidence interval bars represent the conditional marginal effects from a binning estimator at the medians of the three moderator-value bins, the first tercile (L), the second tercile (M), and the third tercile (H). The stacked histogram shows the distribution of the moderator (oil price in U.S. dollars; 2005 = $100), with the red and white shaded bars indicating the distribution of the moderator in the treatment and control groups, respectively.

The model specifications do not include the covariates of casualties or number of fighters because the rhetoric and oil price observations are monthly, whereas the covariate data are yearly. However, we include fixed effects, which our initial model results show produce coefficients in the same direction and at the same level of significance as the specification with the covariates (see Table 3 in the main text and Table S4 in Section D.1). Finally, note that interpreting the independent effect of support origin—and its coefficients, which are sometimes negative—is substantively meaningless because there are no cases in which oil price equals zero (see Brambor, Clark, and Golder 2006).

Part E. EFFECT OF RHETORIC ON FUTURE EXTERNAL SUPPORT

Does adopting a certain rhetoric affect groups' external support in the future? To provide an initial answer, we first calculated the mean rhetoric for each group by year. Aggregating groups' rhetoric into years is necessary because support is observed on a yearly basis. We then use a fractional logistic model, as described in the main text, to analyze, first, the effect of external support on rhetoric and, second, the effect of rhetoric lagged one year on receiving external support. Table S6 presents the results as mean marginal effects.

Table S6. Effect of Lagged Rhetoric on External Support

	(1) Rhetoric ratio	(2) External support
External support	0.134**	
	(0.040)	
Lagged rhetoric		0.305
		(0.576)
Observations	51	41
Pseudo R^2	0.010	0.009

Note: Coefficients are mean marginal effects derived from a fractional logit model. Standard errors, in parentheses, are clustered by group and period.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

External support is positively and significantly associated with the rhetoric ratio score. That is, receiving external support increases groups' rhetoric of subversion. This result aligns with the main findings of our study. However, when we lag rhetoric by a year and estimate its effect on external support, we find no significant relationship. This result offers suggestive evidence that as groups increase their rhetoric of subversion, they do not increase their probability of receiving external support in the following year.

Notes

[1] Version 1.0; https://cran.r-project.org/web/packages/translateR/translateR.pdf
[2] Version 1.3.0; https://cran.r-project.org/web/packages/stm/stm.pdf
[3] Version 0.6.1; https://cran.r-project.org/web/packages/preText/preText.pdf
[4] Although topic models can be repeated, their multi-modal nature could lead to slightly different results (Roberts et al. 2014).
[5] http://ucdp.uu.se
[6] https://icr.ethz.ch/data/epr/acd2epr/
[7] Version 1.0.3; http://yiqingxu.org/software/interaction/RGuide.html
[8] http://www3.nd.edu/ rwilliam/xsoc73994/Margins02.pdf
[9] http://ucdp.uu.se

References

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3:993–1022.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. “Understanding Interaction Models: Improving Empirical Analyses.” *Political Analysis* 14(1):63–82.

Christia, Fotini. 2012. *Alliance Formation in Civil Wars*. New York: Cambridge University Press.

Denny, Matthew, and Arthur Spirling. 2018. “Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, And What to Do About It.” *Political Analysis* 26(2):168–89.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. “Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding.” *Poetics* 41(6):570–606.

Grimmer, Justin, and Brandon Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3):267–97.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. “How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.” *Political Analysis* 27(2):163–92 (https://doi.org/10.1017/pan.2018.46).

Högbladh, Stina, Therése Pettersson, and Lotta Themnér. 2011. “External Support in Armed Conflict 1975–2009, Presenting New Data.” Presented at the 52nd Annual International Studies Association Convention, Montreal, Canada.

Light, Ryan, and Colin Odden. 2017. “Managing the Boundaries of Taste: Culture, Valuation, and Computational Social Science.” *Social Forces* 96(2):877–908.

Lucas, Christopher, Richard Nielsen, Molly Roberts, Brandon Stewart, and Dustin Tingley. 2015. “Computer-Assisted Text Analysis for Comparative Politics.” *Political Analysis* 23(2):254–77.

Mische, Ann. 2014. “Measuring Futures in Action: Projective Grammars in the Rio+20 Debates.” *Theory and Society* 43(3–4):437–64.

Nelson, Laura K. 2017. “Computational Grounded Theory: A Methodological Framework.” *Sociological Methods & Research* (https://doi.org/10.1177/0049124117729703).

Nielsen, Richard. 2017. *Deadly Clerics*. New York: Cambridge University Press.

Quinn, Kevin, Burt Monroe, Michael Colaresi, Michael Crespin, and Dragomir Radev. 2010. “How to Analyze Political Attention with Minimal Assumptions and Costs.” *American Journal of Political Scien*ce 54(1):209–28.

Rana, Muhammad Amir. 2008. “Jihadi Print Media in Pakistan: An Overview.” *Conflict and Peace Studies* 1(1):1–18.

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi. 2016. “A Model of Text for Experimentation in the Social Sciences.” *Journal of the American Statistical Association* 111(515):988–1003.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2017. “stm: R Package for Structural Topic Models.” *Journal of Statistical Software*. DOI:10.18637.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–82.

Roy, Olivier. 1986. *Islam and Resistance in Afghanistan*. New York: Cambridge University Press.

Rubin, Barnett. 2002. *The Fragmentation of Afghanistan*. New Haven, CT: Yale University Press.

Strick van Linschoten, Alex, and Felix Kuehn. 2018. *The Taliban Reader: War, Islam, and Politics*. London, UK: Hurst.

Sundberg, Ralph, Kristine Eck, and Joakim Kreutz. 2012. "Introducing the UCDP Non-state Conflict Dataset." *Journal of Peace Research* 49(2):351–62.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Hunziker Philipp, and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution* 59(7):1327–42.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20:529–44.

Wucherpfennig, Julian, Nils Metternich, Lars-Erik Cederman, and Kristian Skrede Gleditsch. 2012. "Ethnicity, the State, and the Duration of Civil War." *Journal of Conflict Resolution* 64(1):79–115.

Zaeef, Abdul Salam. 2010. *My Life with the Taliban*. New York: Columbia University Press.