

Supplementary Information

Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures

Luwei Ying

Washington University in St. Louis

Jacob M. Montgomery

Washington University in St. Louis

Brandon Stewart

Princeton University

June 2, 2021

The supplementary information to “Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures” includes seven sections. The first section expands on the current validation practices in the literature. The second section shows example HITs for each task as they appear on the online workers’ screen including an example HIT for a non-English text. The third section presents and discusses several “less ideal” validation task designs that we have tried but decided not to recommend for future researchers. This includes an approach to using crowdsourcing to choose between alternative labels. The fourth section summarizes all of our recommendations for validating a topic model and topic labels including practical tips on training and monitoring workers. The fifth section introduces our R package, `validateIt`, and provides a manual walking people through the procedures of posting their own tasks. The sixth section provides an example of the training module we used and discusses workers’ performance. The last section contains additional information on the corpus and model fit.

Contents

1	Current Practices in Political Science: A Survey	SI: 1
2	Example HITs	SI: 6
2.1	Example HITs for Topic Validation	SI: 6
2.2	Example HITs for Label Validation	SI: 7
2.3	Example HITs in non-English (Chinese)	SI: 8
3	Alternative Task Designs and Results	SI: 9
3.1	R4WSI Tasks with Documents	SI: 9
3.2	Two Alternative Label Validation Tasks	SI: 11
3.3	Selecting among Alternative Labels?	SI: 13
4	A Checklist for Crowdsourced Topic Validation	SI: 17
5	Software and Working Example	SI: 19
5.1	R Package: <code>validateIt</code>	SI: 19
5.2	Fitting Candidate Topic Models and Labeling	SI: 19
5.3	Training and Certification	SI: 21
5.4	Preparing Tasks	SI: 23
5.4.1	Preparing Tasks for Topic Validation	SI: 23
5.4.2	Preparing Tasks for Label Validation	SI: 26
5.5	Posting Tasks and Maintaining	SI: 28
5.6	Retrieving and Evaluating Results	SI: 32
6	Workers' Training and Performance	SI: 35
6.1	Training Module for Word Intrusion	SI: 35
6.2	Training Module for Top 8 Word Set Intrusion	SI: 39
6.3	Training Module for Random 4 Word Set Intrusion	SI: 46
6.4	Training Module for Label Intrusion	SI: 50
6.5	Training Module for Optimal Label	SI: 56
6.6	Workers' Performance	SI: 62
7	More on the Corpus, Topic Model Fit, and Labeling	SI: 63
7.1	Word Mass Distribution	SI: 63
7.2	Word Clouds	SI: 64
7.3	Representative Documents for Labeling	SI: 68

1 Current Practices in Political Science: A Survey

To construct the review of recent practice in Section 2.2 of the main paper, we identified all papers published in the *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics* from January 1, 2018 to January 2021 (all articles published online at the time of our search) that included the phrase “topic model.” This resulted in twenty publications. Table SI1-SI3 present the full set of articles and our detailed coding for each of them. Two authors coded each article independently and all three authors discussed cases where there was disagreement to arrive at a consensus coding.

As explained in the main text, we have omitted a number of different types of validations, particularly authors reporting that they have carefully read the documents. While this is doubtless the most important validation strategy, we are explicitly interested here only in (1) validations of topic meaning that are (2) reported in the main body or appendix of the paper, and (3) where the reader can observe the result of the work done. This leaves out a lot of interesting validations but corresponds to our setting of interest.

While coding these articles’ validation practices, we created three dichotomous variables reflecting the most common classes of validation strategies reported (in either the main paper or the appendix only) which met our criteria: topic-specific word lists, fit statistics, and bespoke validation of individual topic meanings. Topic-specific word lists include either the most probable words in the topic under the model or alternative criteria such as frequency and exclusivity (FREX) (Roberts et al., 2016) words. We included both cases where the words were reported in lists or visualized in word clouds. Model fit statistics include held-out log likelihood (Wallach et al., 2009) or surrogate scores such as “semantic coherence” (Mimno et al., 2011; Roberts et al., 2014). These provide a sense of whether or not the model is over-fitting, and some previous research shows that surrogates correlate with human judgements. We also coded the bespoke approaches—additional validations of topic meaning designed especially for their case to establish construct validity. These were more varied and unavoidably subjective. We describe many of the individual procedures below.

In Table SI1-SI3, we also provide more information with regard to the corpus, model choice, and whether or not the measure from topic models is used as a predictor or outcome variable. In addition to the three approaches to validation we coded, several papers took creative and unique approaches to validation that do not fit easily into this coding scheme. We have marked them in Table SI3 and follow the table with short summary notes.

We emphasize that the result of this coding should not be used to critique any given study—articles used topic models for a wide-variety of purposes and to varying degrees as the foundation of their evidence base. For example, Pan and Chen (2018) uses the topic model for exploration (which does not itself need to be validated) and Rozenas and Stukal (2019) uses it as a validation of another measure (where a call for validation could quickly turn into an infinite regress). We estimate in at least 1/4 of the cases, the validations we code arguably aren’t directly relevant because the topic model was not used for measurement or was not a central part of the analysis. However, making this assessment at an article-level proved too difficult to do consistently. Thus we encourage readers to consider these primarily as aggregate statements about the approaches being used in the literature as we have presented them in the main paper.

Table S11: Topic model Usage in Three Political Science Journals

No.	Author Date	Journal	Title
1	Barberá et al. (2019)	APSR	Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data
2	Barnes and Hicks (2018)	AJPS	Making Austerity Popular: The Media and Mass Attitudes toward Fiscal Policy
3	Berliner et al. (2020)	JOP	The Political Logic of Government Disclosure: Evidence from Information Requests in Mexico
4	Blaydes, Grimmer and McQueen (2018)	JOP	Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds
5	Blumenau and Lauderdale (2018)	JOP	Never Let a Good Crisis Go to Waste: Agenda Setting and Legislative Voting in Response to the EU Crisis
6	Dietrich, Hayes and O'Brien (2019)	APSR	Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech
7	Feierherd (forthcoming)	AJPS	Courting Informal Workers: Exclusion, Forbearance, and the Left
8	Horowitz et al. (2019)	JOP	What Makes Foreign Policy Teams Tick: Explaining Variation in Group Performance at Geopolitical Forecasting
9	Jiang and Zeng (2019)	JOP	Countering Capture: Elite Networks and Government Responsiveness in China's Land Market Reform
10	Kim (2018)	APSR	Media Bias against Foreign Firms as a Veiled Trade Barrier: Evidence from Chinese Newspapers
11	Lacombe (2019)	JOP	The Political Weaponization of Gun Owners: The National Rifle Association's Cultivation, Dissemination, and Use of a Group Social Identity
12	Lowande (2018)	JOP	Politicization and Responsiveness in Executive Agencies
13	Magaloni and Rodriguez (2020)	APSR	Institutionalized Police Brutality: Torture, the Militarization of Security, and the Reform of Inquisitorial Criminal Justice in Mexico
14	Martin and McCrain (2019)	APSR	Local News and National Politics
15	Motolinia (2020)	APSR	Electoral Accountability and Particularistic Legislation: Evidence from an Electoral Reform in Mexico
16	Nielsen (2020)	AJPS	Women's Authority in Patriarchal Social Movements: The Case of Female Salafi Preachers
17	Pan and Chen (2018)	APSR	Concealing Corruption: How Chinese Officials Distort Upward Reporting of Online Grievances
18	Parthasarathy, Rao and Palaniswamy (2019)	APSR	Deliberative Democracy in an Unequal World: A Text-As-Data Study of South India's Village Assemblies
19	Roberts, Stewart and Nielsen (2020)	AJPS	Adjusting for Confounding with Text Matching
20	Rozenas and Stukal (2019)	JOP	How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television

Table SI2: Topic model Usage in Three Political Science Journals (Continued)

No.	Corpus	Model	K	Outcome	Predictor
1	Tweets sent by members of Congress, different samples of the public, and media outlets	LDA	100	yes	yes
2	Articles from the <i>Guardian</i> and the <i>Telegraph</i>	STM	6	yes	no
3	Access-to-information (ATI) requests filed with the Mexican federal government agencies	LDA	20	no	yes
4	Medieval Muslim and Christian books	vMF; HTM	4 broad/60 sub	yes	no
5	Legislative summaries from European Parliaments	LDA	29	no	yes
6	Transcripts from the US members of Congress's speeches	STM	30	yes	yes
7	Parcerias (i.e., cooperation agreements) signed by PT mayors and the federal Ministry of Labor in Brazil	STM	5 topics highlighted	no	no
8	Forecast Explanations	STM	45	yes	yes
9	Petitions filed on the Local Leader Message Board in China	LDA	30	no	no
10	News articles on automakers by <i>Beijing Daily</i> , <i>Beijing Youth Daily</i> , <i>Beijing Morning Post</i> , and <i>Beijing Evening News</i>	STM	25	yes	no
11	Editorials of the <i>Rifleman</i>	STM	6	no	no
12	Contact descriptions provided in agency correspondence logs	STM	12	yes	no
13	Mexican jurisprudential theses	STM	5	no	no
14	Transcripts from TV news broadcast	LDA	15	yes	no
15	Transcripts from Mexican legislative sessions	CTM	450	yes	no
16	Text from saaid.net	STM	10	yes	yes
17	Citizen complaints in Online Sentiment Monitoring Reports in China	STM	40	yes	yes
18	Transcribed and translated proceedings of assemblies in Indian villages	STM	15	yes	no
19	Publications (expanded on Maliniak, Powers and Walter 2013) and Weibo posts (Fu, Chan, and Chau 2013)	STM	15 and 100	no	yes
20	Daily news reports from Russia's state-owned television network	STM	80	yes	no

Table SI3: Topic model Usage in Three Political Science Journals (Continued)

No.	Word Distributions	Distri- butions	Fit Statistics	Bespoke Meaning Validation	Bespoke Validation Details	See Notes
1	yes		yes	yes	Custom website	yes
2	yes		no	no		no
3	yes		yes	no		no
4	yes		no	no		yes
5	yes		no	yes	Coded 200 documents whether crisis related	no
6	yes		no	yes	Republicans/Democrats' speeches compared to parti- san topics	yes
7	yes		no	no		no
8	yes		no	no		no
9	yes		no	no		no
10	yes		no	no		no
11	yes		no	no		no
12	yes		yes	no		no
13	yes		no	no		no
14	yes		yes	no		no
15	yes		no	yes	2 topics compared to external events	no
16	yes		no	no		yes
17	yes		no	no		yes
18	yes		no	yes	Predictive validity; Survey of hu- man observers	no
19	yes		no	no		yes
20	no		no	no		yes

Supplemental Notes

- Barberá et al. (2019)
They provide full time-series plots on their custom website for all of their topics that can be used by the reader to compare to important external events. The site also includes sample documents for each topic so that readers can read along themselves. Finally, Barberá et al. (2019) compare the coverage of their topics to key votes in the United States Congress.
- Blaydes et al. (2018)
The replication archive includes paragraph-long descriptions of each topic and sub-topic informed by their reading of the texts.
- Dietrich et al. (2019)
In an analysis of a topic chosen to represent women, Dietrich et al. (2019) compare their results based on the topic model to results based on a dictionary method designed to measure the same concept. By our coding rules this does not validate the meaning of the topic which is why in the table above we focused on the partisan topics analysis which is more ancillary in the original paper.
- Nielsen (2020)
The appendix includes extensive evidence of the robustness of the results to different number of topics. The extensive replication archives also allow for the interested reader to explore more deeply.
- Pan and Chen (2018)
As noted in a footnote of the original paper, topic models are used here primarily as an exploratory device. The eventual evidence is based on a supervised learning method. Thus, the topic model itself isn't directly evaluated for meaning, but it doesn't need to be because it served its purpose.
- Roberts et al. (2020)
This paper presents text matching, where the goal of the topic model is to adjust for confounding by conditioning on the high-dimensional text. Roberts et al. (2020) present approaches to balance checking, conduct a hand-coding comparison of matched pairs and report example matched pairs. These are relevant validations for this particular context, but are not validations of the topic meaning in the sense we are exploring here.
- Rozenas and Stukal (2019)
In this case, the topic model is being used as a validation check on a measure using supervised learning and thus is described more minimally (since to validate the validation would create a problem of infinite regress). In some sense this approximately fits our description of bespoke validation where the 'model is detecting something we are ex ante confident is true' (where the true thing is what is found earlier in the analysis). We ultimately decided not to consider this a case of bespoke validation because the expectation isn't fixed prior to the study. Based on the appendix, our coding for word

distributions was ‘no’, but here it is worth noting that the replication archive (which we did not consider in our coding) contained a list of FREX words for each topic. In total, this was a difficult edge case.

2 Example HITs

This section provides example HITs for The tasks we proposed in the paper, showing one screenshot for each of these tasks from the workers’ view on the Mturk platform.

2.1 Example HITs for Topic Validation

Figure SI1: An Example HIT for Word Intrusion (WI)

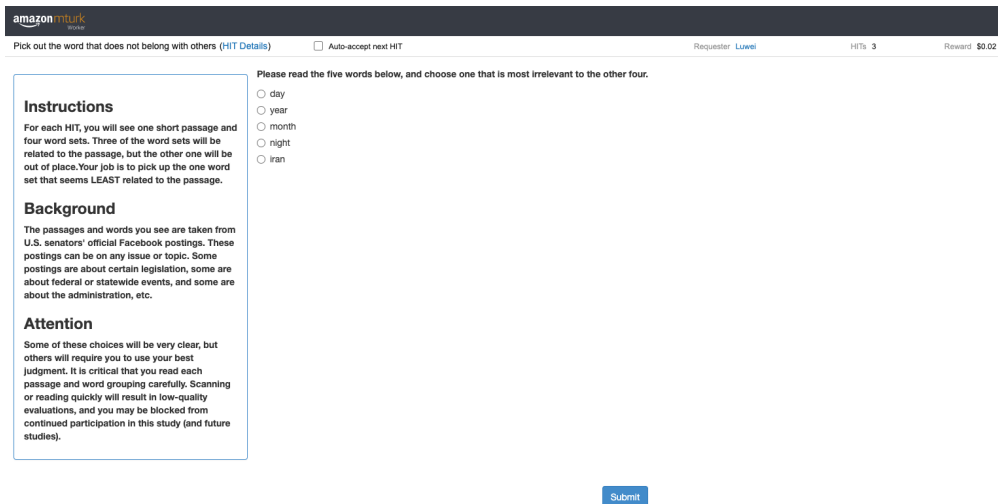


Figure SI2: An Example HIT for Topic 8 Word Set Intrusion (T8WSI)

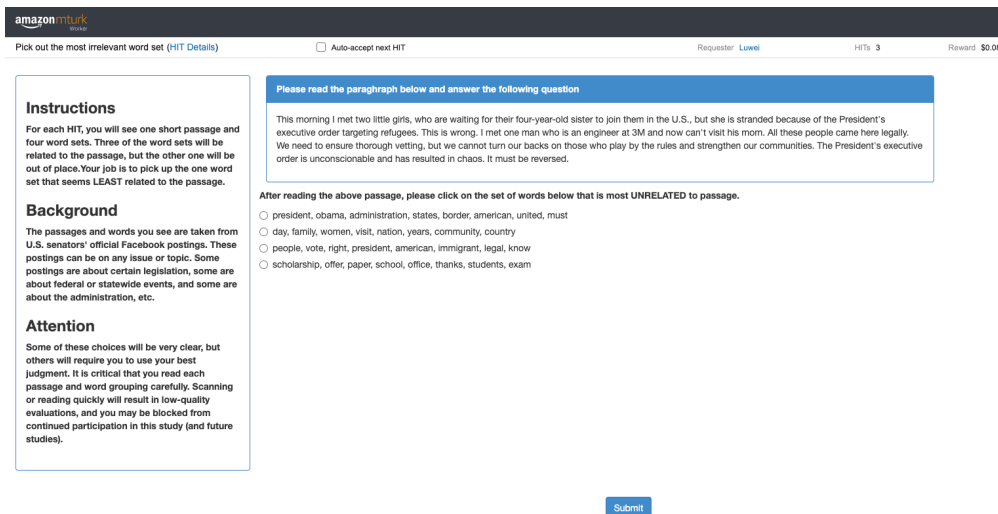


Figure SI3: An Example HIT for Random 4 Word Set Intrusion (R4WSI)

amazon mturk
Pick out the most irrelevant word set (HIT Details) Auto-accept next HIT Requester: Luwel HITs: 3 Reward: \$0.06

Please click on the word set below that is most UNRELATED to the other three.

Attention
Some of these choices will be very clear, but others will require you to use your best judgment. We understand that in many cases it will be hard to tell what the "right" choice is. Use your best judgment, but please be attentive. We have randomly mixed in a number of gold-standard HITs whose answers are less ambiguous to evaluate the quality of your work. If you miss a large number of gold-standard HITs, you may be blocked from continued participation in this study (and future studies).
After you've done a relatively large number of HITs of a specific task structure, your qualification might be temporarily suppressed. This does not mean you've done anything wrong, but we need to ensure that a variety of workers complete our tasks. This is NOT a block and will have NO negative impact on your Mturk record.

Instructions
For each HIT, you will see four sets of words. Three of them will be related to each other, but one word set will be out of place. Your job is to pick up the one word grouping that does NOT belong with the others.

Background
The words you see are taken from U.S. senators' official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

- family, every, work, country
- workers, continue, women, everlasting
- fighting, paid, living, ensure
- technology, ai, google, machine

Submit

2.2 Example HITs for Label Validation

Figure SI4: An Example HIT for Label Intrusion (LI)

amazon mturk
Pick out the most irrelevant label (HIT Details) Auto-accept next HIT Requester: Luwel HITs: 2 Reward: \$0.08

Please read the paragraph below and answer the following question

Small businesses create the majority of new jobs in the U.S. One key way we can help these businesses grow is to give them the resources they need to break into the global marketplace. That's why I was glad to join fellow state leaders yesterday to announce that the \$496,066 Small Business Administration STEP grant will go to a program that does just that - help small businesses increase their exports.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

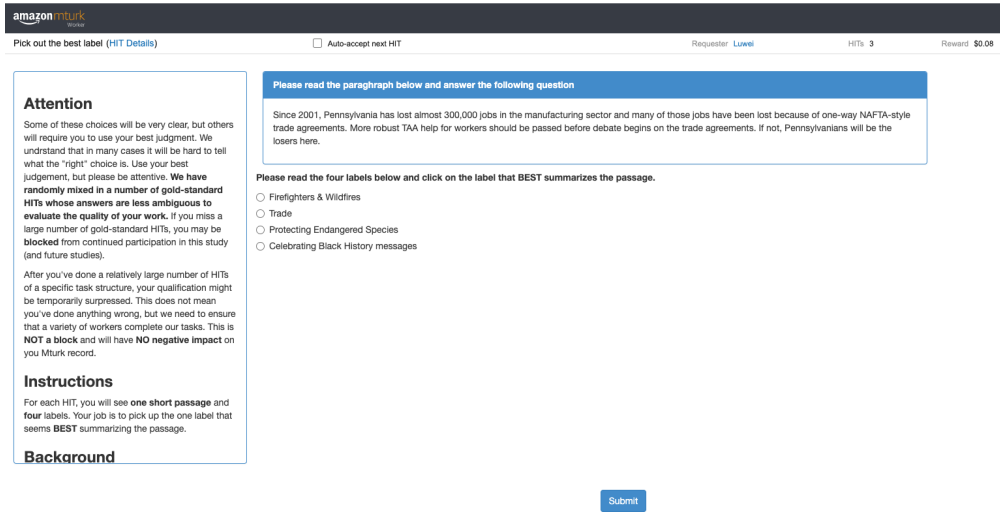
- Resources for Local Communities
- Jobs
- Renewable Energy
- Celebrate Small Businesses messages

Attention
Some of these choices will be very clear, but others will require you to use your best judgment. We understand that in many cases it will be hard to tell what the "right" choice is. Use your best judgment, but please be attentive. We have randomly mixed in a number of gold-standard HITs whose answers are less ambiguous to evaluate the quality of your work. If you miss a large number of gold-standard HITs, you may be blocked from continued participation in this study (and future studies).
After you've done a relatively large number of HITs of a specific task structure, your qualification might be temporarily suppressed. This does not mean you've done anything wrong, but we need to ensure that a variety of workers complete our tasks. This is NOT a block and will have NO negative impact on your Mturk record.

Instructions
For each HIT, you will see one short passage and four labels. Three of the labels will be related to the passage, but the other one will be out of place. Your job is to pick up the one label that seems LEAST related to the passage.

Submit

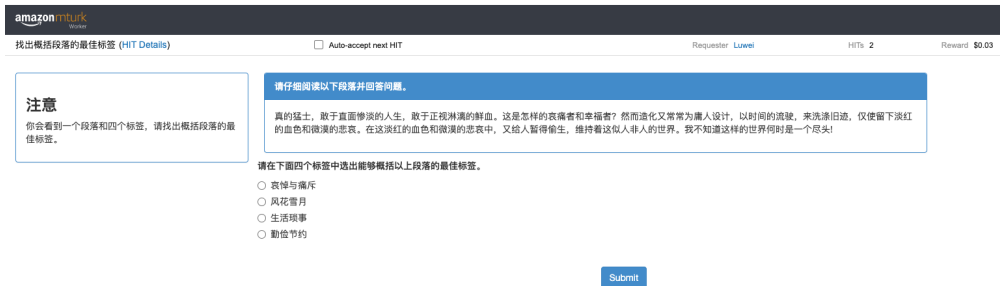
Figure SI5: An Example HIT for Optimal Label (OL)



2.3 Example HITs in non-English (Chinese)

Our package allows researchers to post tasks in different languages with no need to change any settings. Although we didn't discuss in depth posting validation tasks for non-English corpus, we note this possibility here. Figure SI6 shows an example HIT in Chinese, posted with our R package.

Figure SI6: An Example HIT in non-English (Chinese)



3 Alternative Task Designs and Results

In this section, we present two variants of the Random 4 Word Set Intrusion (R4WSI) model validation task and three other task structures for label validation. In each case, we experimented with these approaches but decided not to recommend them as the primary choices for the reasons discussed below. However, for the sake of transparency we report the procedures and results here as they may be of interest to future researchers.

3.1 R4WSI Tasks with Documents

We experimented with two variations of R4WSI task with the major difference that it shows coders a document in addition to the word sets. R4WSI-Random, chooses a random document from the corpus to display. Workers are then asked to identify the word set that does not belong. They must then choose from four different word sets, which are drawn following the same procedure as we did for R4WSI in the main text: the three non-intruder word sets are all chosen from the same topic (the topic most associated with the document) and the intruder set from a different topic.

The second variant is identical to the above R4WSI-Random task with one exception: the documents shown are always one of the ten most representative documents for a specific topic. Thus the only difference between these two variants is how the document is selected. For convenience, we call it R4WSI-Representative. Both variants are summarized in Table SI4.

Table SI4: Additional Task Structures for Coherence Evaluations

	A. Text Presented	B. Question Asked	C. Options Given
R4WSI-Random	A randomly selected document	After reading the above passage, please click on the word set below that is most UNRELATED to the passage.	Three word sets (each containing four mass-based selected words) from the top twenty high-probability words of one topic and one word set (the intruder) mass-based selected from the top twenty high-probability words of another topic
R4WSI-Representative	A document randomly selected from the top ten most representative documents for a given topic	After reading the above passage, please click on the word set below that is most UNRELATED to the passage.	Three word sets (each containing four randomly selected words) from the top twenty high-probability words of one topic and one word set (the intruder) randomly selected from the top twenty high-probability words of another topic

As in the main text, we posted 500 HITs twice, summing to 1000 HITs, for each task-model combination. Note that the R4WSI-Representative task was implemented with four different structural topic models from Models 1-5 in the main text:¹

1. a 20 topic model (same as Model 2) limited to one EM iteration (which keeps the model from properly converging) with random initialization;

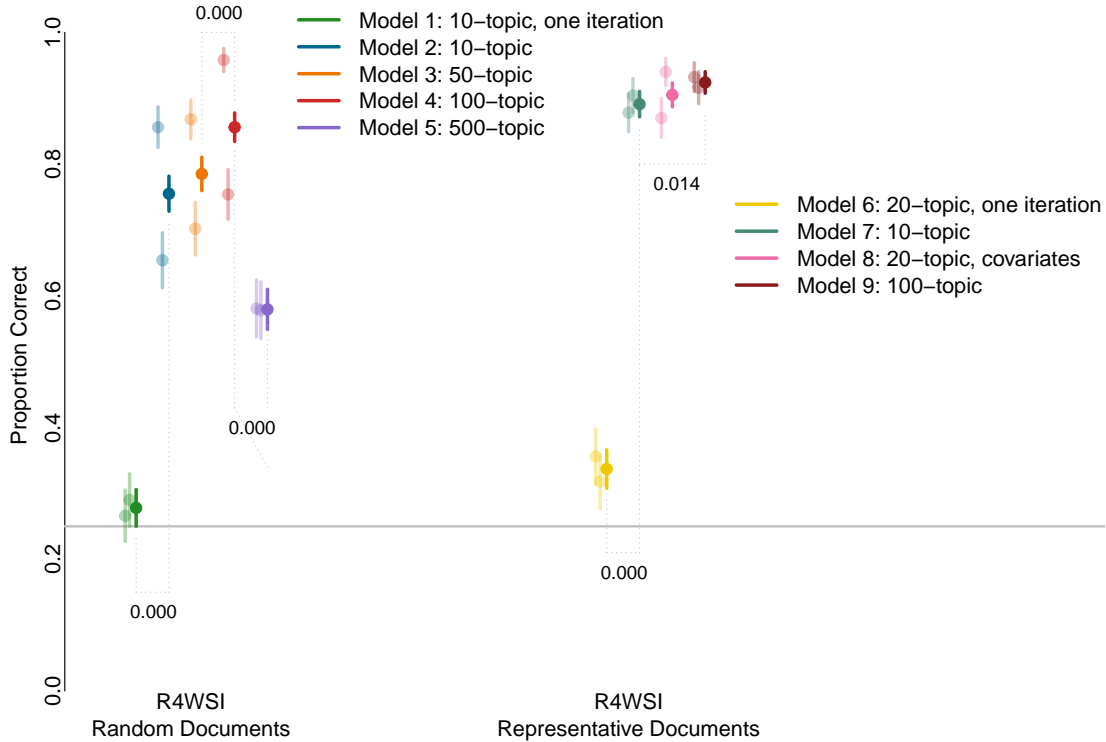
¹Models 6-9 were from an earlier version of the paper where we fit topic models without holding out any documents.

2. a model with 10 topics and no covariates;
3. a model with 20 topics and adding the senators' names as a covariate, and;
4. a model with 100 topics.

The result can be seen in Figure SI7. The main problem is that the R4WSI-Random scores are not reliable across runs for Model 2-4. The scores differed dramatically even though each run includes identical tasks. We were unable to improve the reliability for this task structure even after multiple attempts to recruit high quality workers.

In R4WSI-Representative, the two identical trials for Model 8 are significantly different indicating lower levels of reliability. Additionally, this task structure provided much lower power to distinguish between models. We think this is because the R4WSI-Representative task only uses the clearest documents and thus results in a ceiling effect.

Figure SI7: Results for Coherence Evaluations



Note: The 95% confidence intervals are presented, where the two light bars represent two identical trials (500 HITs each) and the dark bar represents the pooled result (1000 HITs). p -values are based on the pooled set of tasks based on a difference-in-proportions test. No identical trials are significantly different from each other. The grey horizontal line represents the correct rate from random guessing.

3.2 Two Alternative Label Validation Tasks

We have also designed two alternative label validation tasks that rely only on the labels and high probability words. The *Word Set Intrusion* task copies the R4WSI described in the topic validation section in the main text, simply adding the label at the top of the task. An example is shown in Panel (a) of Figure SI8. The *Optimal Label for Word Sets* (OLW) task presents a topic, represented by the eight highest-probability words, and four labels. One label is for the relevant topic and the others are for other topics in the model but within the specified broad category. An example is shown in Panel (b) of Figure SI8. Both task structures are summarized in words in Table SI5.

Table SI5: Additional Task Structures for Label Validation

	A. Text Presented	B. Question Asked	C. Options Given
WSI	A label ^a	Please click on the word set below that is most UNRELATED to the above label.	Three labels for the top three high-probability topics and one label for other economy-related topics
OLW	The eight highest-probability words for a topic	Please read the four labels below and click on the label that BEST summarizes the word set.	One label for the highest-probability topic and three labels for other economy-related topics

^aThe label for one of ten economy-related topics in this empirical illustration.

Past experience teaches that word sets are easier to associate with a topic than an actual document since language out of context is easier to re-interpret. This can make associating broad concepts with word sets easier than with documents, but often at the expense of verisimilitude. Thus, it should not be too surprising to find that validations based primarily on word sets will give less clear guidance.

To test these task structures we followed the same basic procedures as discussed in the main text. For each task/coder combination we created 500 tasks that were coded by trained workers on AMT.² The results from 4,000 high quality HITs are shown in Figure SI9. However, unlike the main text, these task structures were tested with Model 9 described above. In addition, we created “cursory coder” and “careful coder” labels for only 10 economics-related topics (although we followed the same basic procedures as in the main text to create these competing label sets).

The expected result that the “careful coder” labels were superior, however, was not confirmed by the tasks based around word sets (WSI and OWS). Ex ante this makes sense since the “cursory coder” labels were based only on the word sets to begin with. Given that the “careful coder” labels were intentionally designed to more accurately represent the underlying meaning of the topics as they appear in the actual documents, we infer that label validation should be based on documents rather than word sets alone. In all, based on these results we would recommend that labels be evaluated using the LI and OL.

²Workers were paid 0.02/HIT for the OLW task, 0.03/HIT for the WSI task

Figure SI8: Example HITs for Alternative Label Validation Tasks

Resources for Local Communities

Please read the four word sets below and click on the word set that is most **UNRELATED** to the label.

- pushed, make, tools, ensure
- store, started, businesses, like
- communities, tools, continue, make
- possible, reach, needs, pushed

Submit

(a) Word Set Intrusion

today's, times, percent, press, wall, street, unemployment, says

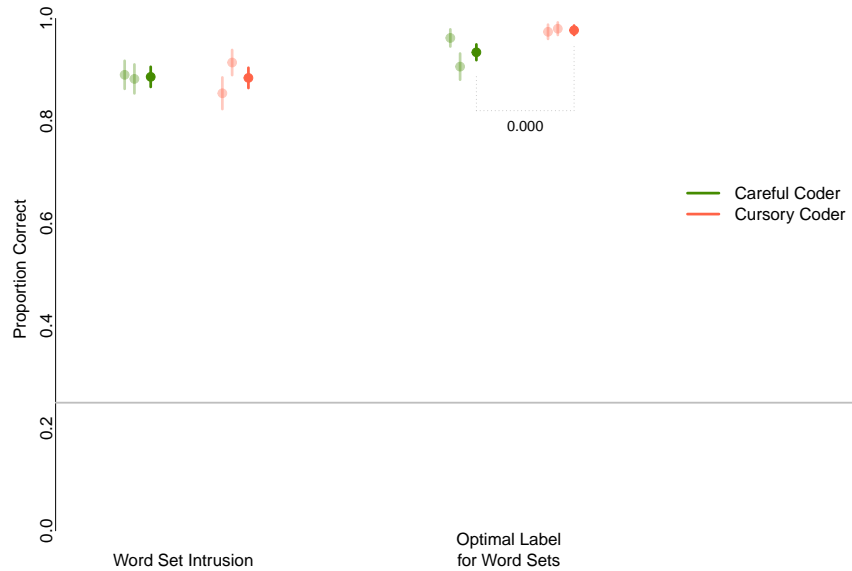
After reading the above word set, please click on the label below that **BEST** summarizes these words.

- Resources for Local Communities
- Income Inequality
- Raising the Debt Ceiling
- Jobs & Keystone Pipeline

Submit

(b) Optimal Label for Word Sets

Figure SI9: Results for Label Validation



Note: The 95% confidence intervals are presented, where the two light bars represent two identical trials (500 HITs each) and the dark bar represents the pooled result (1000 HITs). p -values are based on the pooled set of tasks based on a difference-in-proportions test. No identical trials are significantly different from each other. The grey horizontal line represents the correct rate from random guessing.

3.3 Selecting among Alternative Labels?

A further variation on our labeling task structure focused on testing alternative labels for individual topics (instead of competing label *sets*). Specifically, we presented our good labels (the “careful coder” labels) to two research assistants and asked each of them to come up with three closely related concepts without looking at the top words or documents.³ We then selected three out of these six alternative labels, mixing them with the original “careful coder” labels to construct the tasks. In this way, the alternative labels would, by definition, be similar in meanings. However, we wanted to test the expectation that, *on average*, our

³The instruction we gave the coders are as follows:

Some background information: We are researchers who want to know what the US senators are talking about online. We scraped the senators’ Facebook pages (all posts) from 2008 to 2017 (10 years) and ran a topic model to discover the topics from this corpus. We selected 10 topics and labeled them below. Your job is, by just looking at these labels, to come up with 3 alternative labels for each of the 10 labels. Your alternative labels should be related concepts/phrases that could potentially summarize a similar set of documents. For example, for the label “Terrorism,” three alternative labels could be “Counter-Terrorism,” “Domestic Security”, and “Violent Groups”.

The labels: Equal Pay for Women; Healthcare/Reproductive Rights; Agriculture; Student Loan/Debt; Drug Abuse; Higher Education/Job Training; Wall Street/Financial Sector; Government Shutdown/Congressional Budget; Obamacare/Tax Policy; Deficits/Debt/Budget

“careful coder” labels will outperform alternatives.

To be clear, this alternative task structure is identical to the *Optimal Label* task in the main text. The difference is that coders were not choosing between labels of different topics, but instead choosing directly between *four alternative labels for the same topic*. Like other experiment, we post 500 HITs (50 HITs for each topic) twice. The results (including the four alternatives for each topic) are shown in Figure SI10.

Although the sample size allocated to each topic is small (50 HITs) for each topic, the results are fairly consistent from trial to trial. Our preferred labels perform best in eight out of the ten cases for one run and seven out of the ten cases for a second run. In the two topics (“equal pay for women” and “drug abuse”) where the alternative labels outperform the original labels, revisiting the document sets indicated that these are indeed good labels for those specific topics.

Although these results seem promising, our concern is that the task structure will overly favor specific labels when available leading to confused results. Conceptually, we might imagine having a topic ‘A’ that includes sub-components ‘1’ ‘2’ and ‘3.’ For example, assume a topic model had a broad “economic policy” topic that encapsulated the “government shutdown/congressional budget”, “Obamacare/tax policy”, and “deficits/debt/budget” topics discussed above.⁴ For each document, coders would then choose between: economic policy, government shutdown/congressional budget, Obamacare/tax policy, and deficits/debt/budget. While economic policy might better represent the overall topic, each document will fit in better with the more specific label. Since workers do not have access to the whole document set or other model outputs, they will (incorrectly) tend to choose the more specific labels even though they are inappropriate for the broad topic.

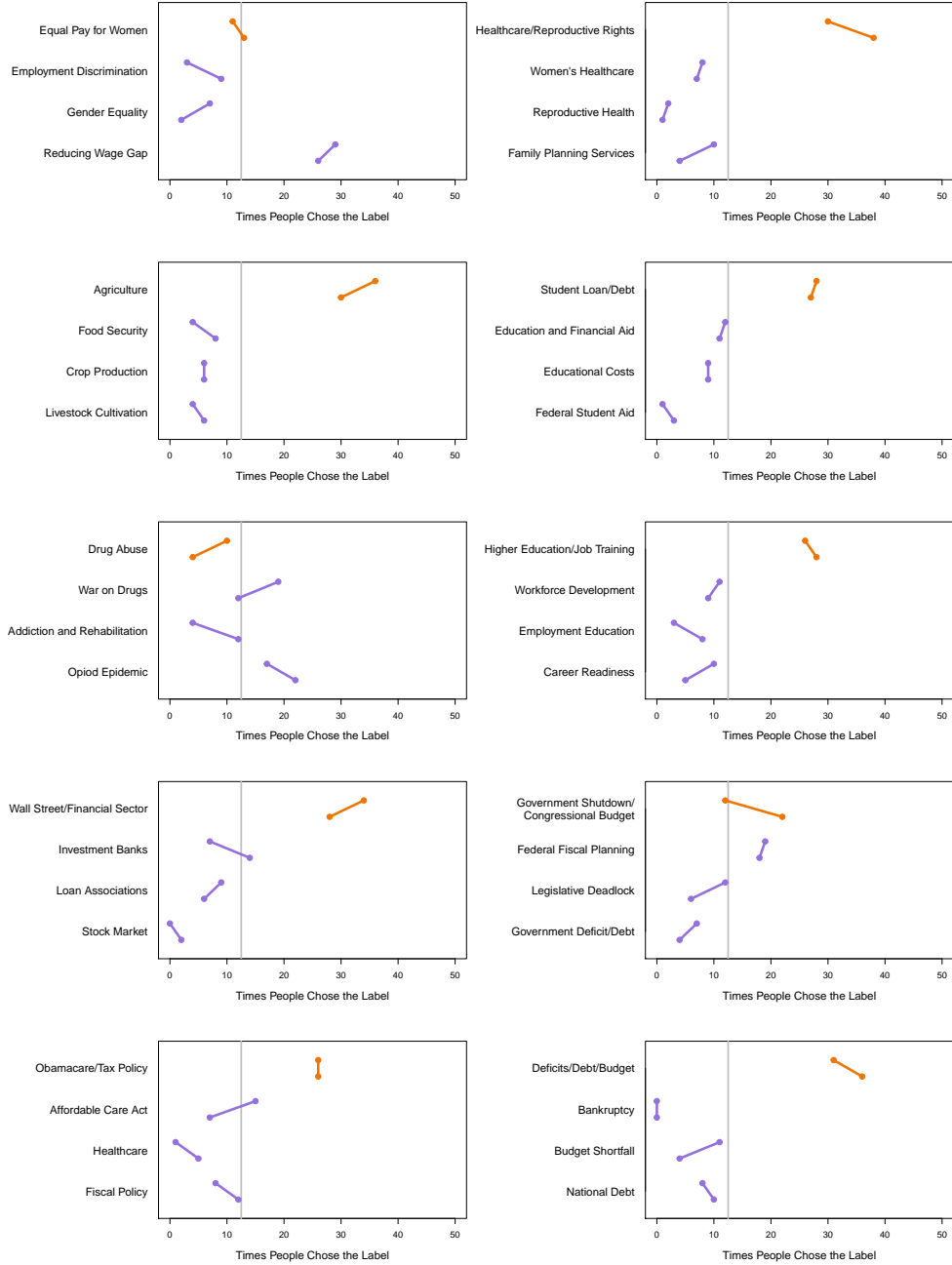
To test this, we re-ran the above experiment but collapsed these three categories and presented respondents with exactly this choice set. The results confirmed our suspicion that coders would strongly prefer the more specific labels. For the 150 relevant tasks, coders chose the broader “economic policy” topic only 20 times using the more specific labels 86% of the time. Meanwhile the other topics were selected nearly an equal proportion of the time (government shutdown/congressional budget was chosen 47 times, deficits/debt/budget chosen 45 times, and Obamacare/tax policy 48 times). A naive analyst looking at these results would conclude that all of the labels are poor, but that the “economic policy” label is worst (even though it is the most appropriate for this setting).

Thus, while this approach may be useful for helping researchers compare competing topic labels, we are hesitant to recommend its widespread adoption since the outputs seem inherently ambiguous. Low scores for a specific label may indicate a poor conceptual mapping but may also simply indicate that the competing topics include sub-components. That is, coders may regularly choose the more specific labels for any given document and never the broader “theme” appropriate for the topic.

Further, in the context of topic models, the goal is not simply to place the “right” label on any given topic but to also ensure that the conceptual category is internally coherent and distinct from neighbors. That is, we are not typically labeling topics in isolation but as a set. The idea is not simply to find the best label for one specific topic, but to find a set of

⁴Collapsing adjacent topics into an overarching topic occurs commonly, for instance, when reducing the number of topics in a model.

Figure SI10: Results for the Alternative Label Validation Task



Note: The “careful coder” labels are in orange and the alternative labels are in purple. The two dots represent two identical trials. The gray vertical lines represent random guessing.

labels that we can clearly distinguish. Thus, this task structure may be useful for testing between competing conceptual labels but only if care is taken to avoid this decomposition fallacy.

4 A Checklist for Crowdsourced Topic Validation

In this section, we provide succinct checklist for future researchers who intend to use topic models to measure concepts. This list is intended to be descriptive of the steps we have found to be useful in practice rather than prescriptive. We doubt there is any finite set of recommendations that will be appropriate in all settings and this should not be read as a narrow set of requirements but rather as a place to start when thinking through these issues. Still it may be useful to provide a more coherent summary of our approach in one location.

When validating topic coherence, here are the steps we follow.

1. Know your corpus and read it in depth. There is no substitute for being familiar with the content of your documents.
2. Fit several models, then use fit statistics and qualitative reading to select one (or several) you wish to consider further. At a minimum this should be done by reading multiple documents for each topic and considering model outputs such as high probability and FREX words. However, in our experience it is easy to get fooled by this alone. Further steps may be to examine model outputs such as fit statistics (e.g., held-out log likelihoods) and surrogate scores.
3. Consider the effects of alternative pre-processing steps, starting values, and tuning parameters. Themes and topics that are highly sensitive to these choices may be harder to uncover reliably with unsupervised methods.
4. Validate the chosen model(s) using the crowdsourced procedure:
 - (a) Create an Amazon requester account. You might consider also setting up a worker account to see your own tasks as a worker.
 - (b) Generate tasks using our package, be sure to look at them in detail and see if you can do the tasks yourself. If not, you may want to try more models or improve/refine training. Look for common issues or areas of confusion and think about how to set up your training module to explain these issues clearly.
 - (c) Design the training module, explain the background to the workers in detail, and make sure the example tasks cover a wide range of your topics. You will want to include several example HITs that illustrate the easier cases but also some that specifically illustrate edge cases or unusual features of your data. The final step is to include a “test” to screen out poor workers or bots and ensure that only qualified workers engage in your tasks. Keep in mind that harder tests may improve worker quality, but may also decrease the pool of workers who can pass it (and may require higher compensation rates).
 - (d) Post tasks with the free Mturk sandbox version to make sure everything looks right before spending money on Mturk.
 - (e) Remember that the quality of the work is vital. Think carefully about the requirements you want workers to meet (we recommend using master workers). As data comes in, pay attention to the work as it arrives. Be aware to abnormal activities (e.g., submitting a task every second, missing a bunch of gold-standard HITs) and stop these workers early. Workers won’t mind so long as you don’t “reject” their work (which can have huge effects on their reputation scores and

- future income). Simply removing their permission to continue work is enough.
- (f) We recommend that 1/10 HITs should be a “gold standard” HIT where the answer is clear and unambiguous. These can be generated in advance and chosen by the researcher. Our software allows you to include these easily while not contaminating your data. We removed permission from any worker who misses more than two gold HITs.
 - (g) Don’t let any one worker complete too much of your work. We recommend allowing any one worker to only complete 20% of your HITs. This makes it easier to diagnose if there is a problem with your model or the specific workers who picked up your HITs.
 - (h) Decide on a threshold in advance where you will simply discard all of the work from a “bad” worker. We discarded work from two workers who missed four gold HIT.
5. We compared models using a simple two-sample difference in proportions test, although we do not think it is always necessary to evaluate more than one model.
 6. Report the results of the validation including mean, standard error, and sample size. Authors might also report their training module (or a summary of the module) and basic information about any issues with worker quality. While there is no absolute minimum standard for what score is “high enough,” the results should clearly be better than random guessing. This should be only one validation reported along with other forms of model validation.

When validating topic labels, here is our advice.

1. Think about the use-case: Are you looking for a broader concept (e.g., politics) that is captured a set of topics? Or by just one topic? Remember that fine-grained distinctions will by their nature be harder to detect, less reliable, and noisier. If high levels of precision/accuracy are crucial to your design, you may want to go with broader themes (clusters of topics) rather than single topics. Try fitting models with a larger number of topics. It may be easier to combine several smaller, coherent topics rather than try and force meaning onto a single large topic.
2. Label the topics by carefully reading the high probability words, frequent and exclusive words, many representative documents, and other relevant information. Preferably, have multiple people do so independently and reach a consensus. Our method is about validating labels, but you need to take time and give careful thought to developing them in the first place.
3. Label the relevant topics both on and off the concept (e.g., about politics and not about politics). However, in our experience some topics will not be interpretable, mixing themes or focusing on unimportant features of language such as style. Do not try and force a meaning onto your topics where they do not exist. You can simply exclude incoherent topics from your further validation (although they should also be excluded from your eventual analysis).
4. Follow the above tips, now making sure the training module captures the concept you intend to measure. Keep your attention on the worker quality.

5. The LI and OL task structures are both appropriate. The OL task is better suited for single-membership models or where you tend to think of documents as mostly being in one conceptual bin. Where the mixture of topics is more important, the LI task will be more suitable.
6. Report the results with other forms of validation discussed in the main text. You can show the overall score, but may want to focus on the “within” and “between” results when using the broader categorization approach discussed in the main text. The reported results should include the mean score, standard error, and sample size. Authors might also report their training module (or a summary of the module) and basic information about any issues with worker quality.

5 Software and Working Example

Along with the paper, we provide free-to-use software to facilitate the topic and label validation procedures for future researchers. This section serves as a user manual.⁵

5.1 R Package: `validateIt`

The R package, `validateIt`, to implement the methods we proposed in the paper is currently being prepared for submission to the *Comprehensive R Archive Network* (CRAN). Users can install the most recent development version from `GitHub` using the `devtools` package. First, users would have to install `devtools` using the following code. Note that this step only has to be done once:

```
if(!require(devtools)) install.packages("devtools")
```

Then you can load the package and use the function `install_github`

```
library(devtools)
install_github("Luwei-Ying/validateIt", dependencies = TRUE)
library(validateIt)
```

Note that this will install all the packages suggested and required to run our package. It may take a few minutes the first time, but this only needs to be done on the first use. In the future users can update to the most recent development version using the same code.

5.2 Fitting Candidate Topic Models and Labeling

From here, we provide example code to post validation tasks. It shows the entire procedure of validating topic model outcomes using our proposed methods. This procedure consists five major steps: 1) Fitting Candidate Topic Models and Labeling; 2) Training and Certification; 3) Preparing Tasks Locally; 4) Posting Tasks to Mturk and Maintaining Them Online; and 5) Retrieving and Evaluating Results.

⁵This section features simplified instructions for using the proposed method. It is for illustration purposes. To view the original files and code for our study, please refer to the replication materials.

Researchers need to fit several candidate topic models to be validated later. To illustrate the methods, we fit 5 topic models using the R packages `stm` with 500 topics, 100 topics, 50 topics, 10 topics, 10 topics but restricted to only 1 EM iteration, as described in the main manuscript. Topic models generated by other software will also work as long as the users keep track of a) the documents in the topic model; b) the vocabulary in the topic model; c) the matrix of topic conditional word probabilities, usually known as the beta matrix; and d) the matrix of the topic distribution for each document, usually known as the theta matrix.

While this section does not intend to provide detailed instructions on fitting topic models, we want to point out that we did not stem the words. We also randomly select 10% of the documents (16364 out of 163642 documents) and **held out** 50% of the tokens in these documents so we will be able to compare the results from our methods with held-out likelihood. Code for this procedure is included below.

```
library(stm)
# step 1: pre-process, obtain the stm object
# define the customized stop words outside of the function
customstopwords <- c("senator", "senate", "sen", "facebook", "please", "today",
                    "will", tolower(state.name), tolower(state.abb), "rhode",
                    "hampshire", "jersey", "york", "carolina", "dakota")
docs = textProcessor(documents = as.character(Senate$text),
                    metadata = Senate,
                    lowercase = TRUE,
                    removestopwords = TRUE,
                    removenumbers = TRUE,
                    removepunctuation = TRUE,
                    stem = FALSE,
                    wordLengths = c(3, 20),
                    language = "en",
                    striphtml = TRUE,
                    customstopwords = customstopwords)
stmPrep <- prepDocuments(docs[[1]], docs[[2]], meta = docs$meta)
save(stmPrep, file = "Corpus/stmPrep.Rdata")

# step 2: heldout some words from some documents
heldout <- make.heldout(stmPrep$documents, stmPrep$vocab, seed = 123)
save(heldout, file = "Corpus/heldout.Rdata")
documents <- heldout$documents
vocab <- heldout$vocab

# step 3: fit stm with different topic numbers, i.e., different k
stm <- stm(documents, vocab, k = 10)
save(stm, file = "Models/unstemmed/stm.Rdata")
```

To label the topics, we strongly recommend users to follow the practice of the “Careful Coder”: carefully read the high-probability words and frequent & exclusive words (FREX),

as well as fifty representative documents per topic. Ideally, future researchers will have more than one coder to independently label the topics so they can assess inter-coder reliability.

5.3 Training and Certification

First of all, researchers need to register a Mturk requester account here: <https://www.mturk.com/>.

For each of the task structures, Word Intrusion (WI), Top 8 Word Set Intrusion (T8WSI), Random 4 Word Set Intrusion (R4WSI), Label Intrusion (LI), and Optimal Label (OL), researchers need to make a training module to introduce the background and get online workers familiar with the tasks. While doing so, we provide five practice questions, where people's answers are NOT scored, then eight test questions. The workers need to get 7 or more out of the 8 questions correct to receive the qualification.

The training modules are written in .xml files, each with a separate .xml file coding the answers. Section 6.1 - 6.5 present the training modules used for this paper. For .xml formatting, see the example files "Question.xml" and "Answer.xml" in our replication package.

We then create the qualifications on Mturk:

1) Specify the Mturk options.

```
library(pyMTurkR)
Sys.setenv(AWS_ACCESS_KEY_ID = AWS_ACCESS_KEY_ID)
Sys.setenv(AWS_SECRET_ACCESS_KEY = AWS_SECRET_ACCESS_KEY)

# change sandbox = F when ready to run on MTurk
options(pyMTurkR.sandbox = T)

# use this to test that the pyMTurkR settings are correct
AccountBalance()
```

2) Read in the .xml file for questions and answers.

```
TestQuestions <- paste0(readLines("T8WSIQuestion.xml", warn = FALSE),
                        collapse = "")
TestKey <- paste0(readLines("T8WSIAnswer.xml", warn = FALSE),
                 collapse = "")
```

3) Create the qualification.

```
T8WSIQual <- CreateQualificationType(
  name = "top eight word set intrusion qualification",
  description = "Qualification for "top eight word
               set intrusion" tasks.",
  status = "Active", # allows qual to remain active for users
  test = TestQuestions, # pass questions for test
  test.duration = 60 * 60, # test duration, in seconds
  retry.delay = NULL, # how long until worker can retry test;
                  NULL means never
```

answerkey = TestKey)

The created qualification can be seen on the requester's dashboard under the "Manage" tab after a few minutes.

Figure SI11: Check the qualification, which should appear on the dashboard

amazonmturk Requester Create Manage Developer

Results Workers **Qualification Types**

Manage Qualification Types

Below is a list of your Qualification Types and the corresponding number of Workers.

Create New Qualification Type

Qualification Types					
	Name ▼	ID	Workers who have this Qualification	Creation Date	Description
<input type="radio"/>	word intrusion...	37RZXPVRUD2VWDVXW0BPLJW41931LZ	0	February 4, 2020	Qualification for "word intrusion" tasks. For each HIT, you will see FIVE words. Four of them will be related to each other, but one word will be out of place. Your job is to pick up the one word that does NOT belong with the others.
<input type="radio"/>	top eight word...	3CGA0BEV5XKGUKBSUYGDXXWW5EUK6P0	0	February 4, 2020	Qualification for "top eight word set intrusion" tasks. For each HIT, you will see ONE short passage and four word sets. Three of the word sets will be related to the passage, but the other one will be out of place. Your job is to pick up the one word set that seems LEAST related to the passage.
<input type="radio"/>	random four wor...	385F8X38SRVVPOAF0H7P7EDUWHX0LV	0	January 31, 2020	Qualification for "random four word set intrusion" tasks. For each HIT, you will see ONE short passage and four word sets. Three of the word sets will be related to the passage, but the other one will be out of place. Your job is to pick up the one word set that seems LEAST related to the passage.

4) Save the qualification in case changes need to be made in the future.

```
save.image("Qualifications.RData") # you may need to set file path
```

5) Update a qualification if needed, e.g., fix a typo.

```
load("Qualifications.RData")
T8WSIQual <- UpdateQualificationType(
  qual = T8WSIQual$QualificationTypeId, # keep same qualification id
  description = T8WSIQual$Description,
  status = "Active", # manually set status
  test = TestQuestions, # update test
  test.duration = 60 * 60, # manually set test duration
  retry.delay = NULL, # manually set retry delay
  answerkey = TestKey) # update answer key
```

5.4 Preparing Tasks

Researcher will prepare tasks locally before sending them to Mturk.

5.4.1 Preparing Tasks for Topic Validation

For topic validation, our package, `validateIt`, requires a) the documents in the topic model (`docs`); b) the vocabulary in the topic model (`vocab`); c) the beta matrix in the topic model (`beta`); and d) the theta matrix in the topic model (`theta`).

In the main manuscript, we fit *unstemmed* models. The `combMass()` function takes in a `stm` output and combine mass for words with the same root. The output, `newMass`, is a list of two: the combined vocabulary matrix where words are completed to the most frequent form in that specific topic (`newvocab`) and the combined beta matrix in correspondance with the combined vocabulary (`newbeta`). This step is not required in using our crowd-sourcing methods. We see this procedure as analogous to fitting a stemmed topic model but retain the complete form of words.

```
newMass <- combMass(stm)
```

Now, researchers start formatting the tasks:

1) The `validateTopic()` function creates tasks of the desired type and number. Below is the example code for T8WSI, the structure of which task is the most complicated. We have left out the documents where a portion of words have been held out. The default threshold is 20, meaning that our algorithm draws from the pool of top 20 highest probability words.

```
T8WSItasks <- validateTopic(type = "T8WSI",
  n = 500,
  text = stmPrep$meta$post_text[-heldout$missing$index],
  vocab = newMass[[1]],
  beta = newMass[[2]],
  theta = stm50k$theta[-heldout$missing$index,],
  thres = 20)
```

For WI and R4WSI tasks, leave out the `docs` and `theta` arguments. The default threshold is still 20.

```
# WI
WItasks <- validateTopic(type = "WI",
                          n = 500,
                          vocab = newMass[[1]],
                          beta = newMass[[2]],
                          thres = 20)

# R4WSI
R4WSItasks <- validateTopic(type = "R4WSI",
                             n = 500,
                             vocab = newMass[[1]],
                             beta = newMass[[2]],
                             thres = 20)
```

All word-drawing processes are probability-based (mass-based). Specifically, in preparing WI tasks, the function a) orders the word probability ($\beta[k,]$) for each topic k , then b) randomly draws 4 words from top 1 to `thres` words **based on their corresponding probabilities in k** , and then c) randomly draws 1 intruder word from top 1 to `thres` words from another topic, $\neg k$ **based on its corresponding probability in $\neg k$** .

In preparing R4WSI, the function a) for each topic k , randomly draws 12 words from top 1 to `thred` words **based on their corresponding probabilities in k** and randomly assign them to 3 different word sets, then b) randomly draws 4 intruder words from top 1 to `thred` words from another topic, $\neg k$ **based on their corresponding probabilities in $\neg k$** .

In preparing T8WSI, the function a) uses `vocab` and `beta` to find the top 8 words for each topic k , then b) randomly sample a document from the document pool (`docs`) and calculate the top 3 high probability topics associated with that document using `theta`, and then c) sample an intruder topic from other lower probability topics. Notice that the word drawing procedure here is not probability-based as the function is always looking for the top 8 words regardless of their mass.

The output while specifying “T8WSI” looks as below, where each row represents a task. Each task contains an indicator of the topic (`topic`), a randomly drawn document associated with that topic (`doc`), three non-intruder word sets (`opt2 - opt3`), and an intruder word set (`optcrt`).

2) In addition to generating tasks for validation, we suggest preparing a series of gold-standard HITs to monitor the quality of the works in the future. These gold-standard HITs are the “easy” tasks whose answers are unambiguous.

While preparing gold-standard HITs for the current paper, we select 50 HITs for each task structure from pilot runs on Mturk where two different workers have agreed on the answers in two identical rounds. These selection processes were ****not random**** and we also modified the documents and words to make the answers even clearer. Future researchers could simply generate more tasks than they need and hand pick the gold-standard HITs

Figure SI12: Output of validateTopic()

	topic	doc	opt1	opt2	opt3	optcrt
1	1, 5, 8	Courage and a love of liberty hav...	day, family, wo...	president, obam...	people, vote, rig...	health, care, veterans,
2	10, 3, 1	Watch Senator Leahy announce t...	work, bill, act, n...	discuss, watch, ...	day, family, wo...	people, vote, right, pri
3	5, 8, 1	I congratulate the Argentine peo...	president, obam...	people, vote, rig...	day, family, wo...	discuss, watch, news, .
4	3, 5, 10	Our latest newsletter went out to...	discuss, watch, ...	president, obam...	work, bill, act, n...	people, vote, right, pri
5	1, 3, 2	When I studied for a year at the U...	day, family, wo...	discuss, watch, ...	great, visit, mee...	work, bill, act, need, h
6	1, 2, 10	Events across North Carolina hav...	day, family, wo...	great, visit, mee...	work, bill, act, n...	jobs, businesses, ener
7	1, 8, 10	Like many others, I still remembe...	day, family, wo...	people, vote, rig...	work, bill, act, n...	discuss, watch, news, .
8	10, 4, 8	The St. Louis Post-Dispatch & ST...	work, bill, act, n...	national, service...	people, vote, rig...	jobs, businesses, ener
9	2, 1, 4	I would like to announce the serv...	great, visit, mee...	day, family, wo...	national, service...	people, vote, right, pri
10	9, 6, 1	Social Security and Veterans bene...	health, care, vet...	tax, budget, gov...	day, family, wo...	president, obama, adn

they deem as clear from the excessive tasks. An example of T8WSI gold-standard HITs reads like this:

The Keystone XL pipeline represents not only thousands of jobs and growth for the nation's economy, but also a big step toward American energy independence. We can become energy independent in America within five to seven years, but we must commit to moving forward with important projects like the Keystone XL pipeline.

After reading the above passage, please click on the set of words below that is most unrelated to the passage.

- o jobs, business, energy, new, economy, create, state, economic
- o work, project, forward, need, american, legislation, support, make
- o oil, energy, security, pipeline, administration, states, strategy, must
- o day, family, holiday, summer, beach, play, sunshine, vacation

3) Randomly mix in the gold-standard HITs. The `mixGold()` function ensures that one gold-standard HIT would show up every $\frac{\# \text{ total tasks}}{\# \text{ gold standard tasks}}$ number of tasks. It also assigns a unique id for each of the task.

```
goldT8WSI <- read.csv("goldT8WSI.csv", stringsAsFactors = FALSE)
allT8WSItasks <- mixGold(tasks = T8WSItasks, golds = goldT8WSI)
```

The output looks as below, where the topic column for gold-standard HITs indicates “gold.”

4) Next, record the prepared tasks to a specified path. While doing so, the function will create a list of two, where the first element is the above data frame of prepared tasks and the second element only keeps the documents and word sets (in the case of WI only keeps the words and in the case of R4WSI only keeps the word sets) with randomized order, leaving out other meta data, e.g., the ids. Users can assign the record to an object for immediate use. They can always load it later as well.

Figure SI13: After mixing in gold-standard HITs

	topic	doc	opt1	opt2	opt3	optcrt	id
1	1, 5, 8	Courage and a love o...	day, family, wo...	president, obam...	people, vote, rig...	health, care, vet...	1
2	10, 3, 1	Watch Senator Leahy ...	work, bill, act, n...	discuss, watch, ...	day, family, wo...	people, vote, rig...	2
3	5, 8, 1	I congratulate the Ar...	president, obam...	people, vote, rig...	day, family, wo...	discuss, watch, ...	3
4	gold	I have substantial co...	health-care, car...	budget, govern...	people, right, pr...	overseas, iraq, s...	4
5	3, 5, 10	Our latest newsletter ...	discuss, watch, ...	president, obam...	work, bill, act, n...	people, vote, rig...	5

```
record <- recordTasks(type = "T8WSI", tasks = allT8WSItasks,
                      path = "T8WSI/record.Rdata")
```

5.4.2 Preparing Tasks for Label Validation

The procedure for preparing label validation tasks locally is almost identical to that for preparing topic validation tasks. One extra step is that users need to predict the high probability topics for their documents and define a document pool, from which they would like to draw documents. Typically, the document pool for Label Intrusion (LI) tasks contains all documents whose top three high probability topics are among the pre-defined category (in our example, domestic topics). The document pool for Optimal Label (OL) tasks, on the other hand, contains all documents whose top one high probability topic is among the pre-defined category (in our example, domestic topics). The document pool file – like the one depicted in Figure SI14 – must contain a document column and one or three top column(s) named as “top1” (and “top2”/“top3”).

Figure SI14: Structure of the Document Pool Document

	post_text	top1	top2	top3
1	Tomorrow once again we will vote to protect women's...	9	86	45
2	Yesterday I grilled U.S. financial watchdogs on the col...	65	67	83
3	The payroll tax cut expires in two weeks. Letting it ex...	35	88	89
4	Today we passed the payroll tax cut extension so that...	88	9	22
5	Here's an excellent editorial from the Pioneer Press ab...	62	83	56
6	Ive been fighting to help MN farmers and ranchers rec...	65	94	88
7	For years Ive been pushing the big phone companies ...	65	16	26
8	Today our bipartisan working group in the Senate ann...	44	89	33
9	Today I was in St. Paul to highlight public-private part...	88	46	98
10	Today I held a workforce development forum at Dunw...	62	83	56

Now the users will use the function `validateLabel()`, specifying either `type = "LI"` or

`type = "OL"`. `text.predict` specifies the document pool. `text.name` specifies the name of the document column of the document pool data frame. `labels` are the labels given by human coders, in the same order with `labels.index`. Users could choose to add additional intruder labels through the `labels.add` argument. In our example for illustration, we validate 10 domestic labels while adding 10 international labels as cross-category intruders.

```
documentpool <- read.csv("documentpool.csv", stringsAsFactors = FALSE)
# "documentpool.csv" is equivalent to "top1domText.csv" in our replication materials.
OLtasks <- validateLabel(type = "OL",
  n = 500,
  text.predict = documentpool,
  text.name = "post_text",
  labels = c("Equal Pay for Women",
    "Healthcare/Reproductive Rights",
    "Agriculture",
    "Student Loan/Debt",
    "Drug Abuse",
    "Higher Education/Job Training",
    "Wall Street/Financial Sector",
    "Government Shutdown/Congressional Budget",
    "Obamacare/Tax Policy",
    "Deficits/Debt/Budget"),
  labels.index = c(1, 9, 21, 35, 44, 62, 65, 70, 88, 89),
  labels.add = c("International Trade",
    "Praising Active Military/Military Units",
    "Terrorism",
    "Military Sexual Assault",
    "Nuclear Deterrence/International Security",
    "Air Force",
    "Honoring Specific Veterans",
    "Honoring Veterans/Heroes",
    "Military Operations/Armed Conflicts",
    "Veterans Affairs/Veterans Healthcare"))
```

Like the topic intrusion tasks, we have randomly mixed 50 gold-standard HITs into 500 tasks for both task structure. We recommend future researchers to adopt this “gold-standard HITs” approach as well. The `recordTasks()` function records the tasks at a specified local directory. The two elements in the output list of `recordTasks()` is shown in Figure SI15 and SI16.

Figure SI15: Local Record with Identifiers (`recordTasks()` Output 1)

topic	doc	opt1	opt2	opt3	optcrt	id
1	Americans deserve a #fairsho...	Deficits/De...	Government...	Agriculture	Equal Pay fo...	1
2	Did you know the Republican...	Student Loa...	Higher Educ...	Government...	Healthcare/...	2
3	ICYMI read my op-ed in the l...	Deficits/De...	Healthcare/...	Equal Pay fo...	Agriculture	3
4	Congress has a responsibility...	Equal Pay fo...	Drug Abuse	Agriculture	Student Loa...	4
5	The time is NOW to address t...	Government...	Equal Pay fo...	Student Loa...	Drug Abuse	5
6	The Crofton girl wrote to Pre...	Equal Pay fo...	Obamacare/...	Healthcare/...	Higher Educ...	6
7	Read my op-ed in the Wall St...	Student Loa...	Drug Abuse	Deficits/De...	Wall Street/...	7
8	New House Ukraine bill show...	Agriculture	Student Loa...	Drug Abuse	Government...	8
9	I remain committed to repeali...	Student Loa...	Equal Pay fo...	Deficits/De...	Obamacare/...	9
10	gold	Yesterday, the Senate voted t...	Happy Moth...	Linking to S...	Condolence...	Trade
11	89	Admiral Mullen, Chairman of ...	Equal Pay fo...	Drug Abuse	Wall Street/...	Deficits/De...

Figure SI16: Local Record, Randomized Order, i.e., (`recordTasks()` Output 2)

passage	word1	word2	word3	word4
Americans deserve a #fairshot ...	Deficits/Debt/B...	Equal Pay for Wo...	Agriculture	Government Sut...
Did you know the Republican bi...	Healthcare/Repr...	Government Sut...	Higher Educatio...	Student Loan/De...
ICYMI read my op-ed in the In...	Agriculture	Deficits/Debt/Bu...	Healthcare/Repr...	Equal Pay for Wo...
Congress has a responsibility t...	Agriculture	Student Loan/Debt	Drug Abuse	Equal Pay for Wo...
The time is NOW to address thi...	Student Loan/De...	Government Sut...	Equal Pay for Wo...	Drug Abuse
The Crofton girl wrote to Presid...	Higher Educatio...	Healthcare/Repr...	Obamacare/Tax...	Equal Pay for Wo...
Read my op-ed in the Wall Stre...	Deficits/Debt/B...	Wall Street/Finan...	Student Loan/D...	Drug Abuse
New House Ukraine bill shows ...	Student Loan/De...	Agriculture	Government Sut...	Drug Abuse
I remain committed to repealin...	Equal Pay for Wo...	Student Loan/Debt	Deficits/Debt/B...	Obamacare/Tax ...
Yesterday, the Senate voted to ...	Linking to Speec...	Trade	Condolences me...	Happy Mother's ...
Admiral Mullen, Chairman of th...	Equal Pay for Wo...	Drug Abuse	Wall Street/Fina...	Deficits/Debt/B...

5.5 Posting Tasks and Maintaining

Now, researchers should be able to interact with Mturk and post tasks. They will do so by first manually specifying the basic tasks settings and then sending tasks through the API.

1) Login to the Mturk requester page, click “New Project” under the “Create” tab. Click the “Sentiment Analysis” tab (which allows us to specify our customized layout later) on the left. Then, click “Create Project.”

2) Following the prompts, specify the properties. In particular, be sure to add an additional qualification created in section “Training and Certification” under the “Worker requirements” tab. Our tasks in the paper require a “master” qualification as well. Then, click “Design Layout.”

Figure SI17: Create a Project: Step One

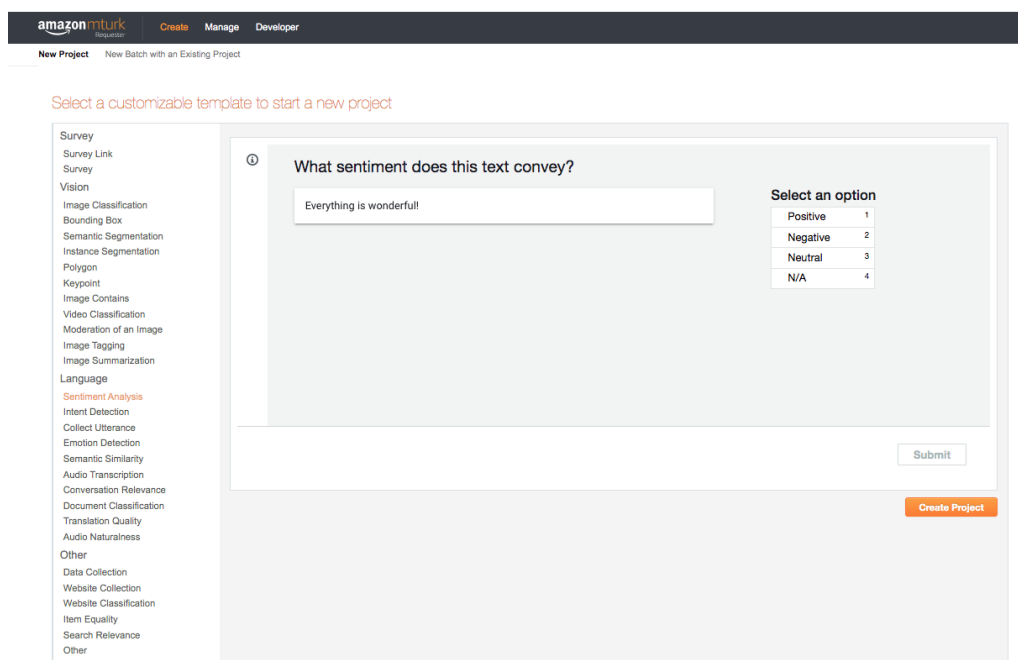
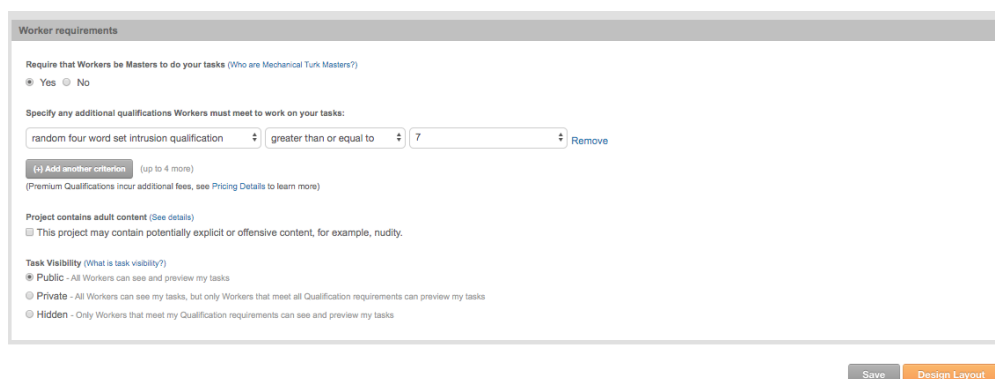


Figure SI18: Create a Project: Step Two



3) Replace the default layout file with our customized layout. The example .html file can be found as a separate file in our replication package (Layout.html). Notice that researchers can change the instructions for their own tasks, which will later appear on the left of the workers' screen. For the tasks in our paper, we specifically tell them:

Some of these choices will be very clear, but others will require you to use your best judgment. We understand that in many cases it will be hard to tell what the “right” choice is. Use your best judgement, but please be attentive. **We have randomly mixed in a number of gold-standard HITs whose answers are less ambiguous to evaluate the quality of your work.** If you miss a large number of gold-standard HITs, you may be **blocked** from continued participation in this study (and future studies).

After you've done a relatively large number of HITs of a specific task structure, your qualification might be temporarily suppressed. This does not mean you've done anything wrong, but we need to ensure that a variety of workers complete our tasks. This is **NOT a block** and will have **NO negative impact** on you Mturk record.

4) Preview the Layout and finish creating tasks.

From here, we can do everything through the API, sending instructions from R to interact with the Mturk platform.

5) Load the prepared tasks from section 5.4.

```
load("T8WSI/record.Rdata")
```

6) Send tasks to Mturk using the `sendTasks()` function. Users must specify a path to **record the HITids immediately** as they are vital for future procedures. The function will record the HITids in a list format where the first element is a vector of HIT ids returned by Mturk and the second element is **the mapping of the local HIT ids to the Mturk HIT ids**. The `tasksids` argument allows users to post specific tasks by specifying the task ids (numeric form) in a vector. If `tasksids` left unspecified, all tasks in the record will be posted.

```
HITids <- sendTasks(hit_type = "FIND_IT_FROM_DASHBOARD",  
                  hit_layout = "FIND_IT_FROM_DASHBOARD",  
                  type = "T8WSI",  
                  tasksrecord = record,  
                  tasksids = c(1:10, 15),  
                  HITidspath = "T8WSI/HITids/testIDs.Rdata")
```

```
# Console responds  
> Sending task to MTurk  
> HITids saved to T8WSI/HITids/testIDs.Rdata
```

Notice that `hit_type` and `hit_layout` can be found from the MTurk requester's dashboard by clicking the project name, e.g., "T8WSI":

Once posted, worker should be able to see the tasks on the Mturk platform. They can get the certification by going through the training module and passing the test. Figure SI20 depicts an example task from the workers' perspective, we see that a "master" qualification is also required.

Figure SI19: Finding HITType ID and Layout ID

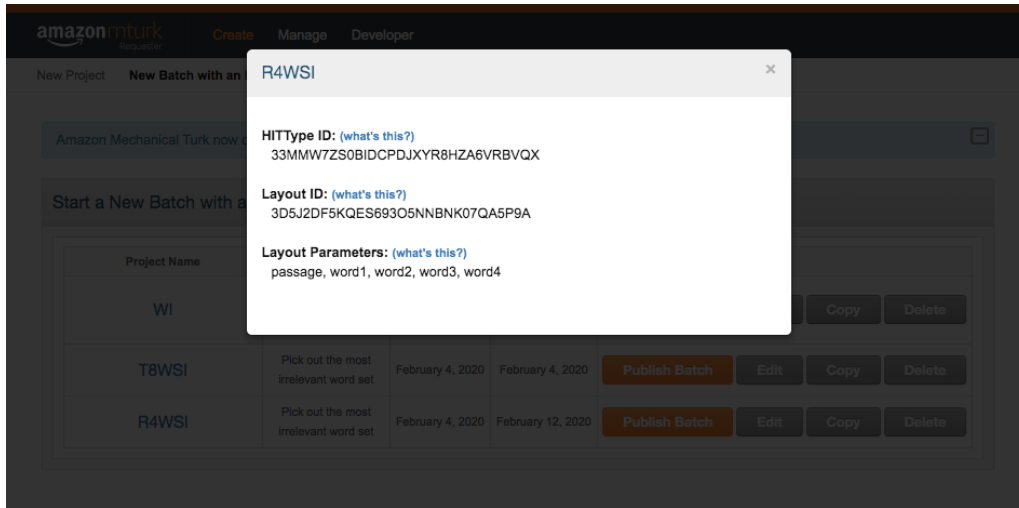
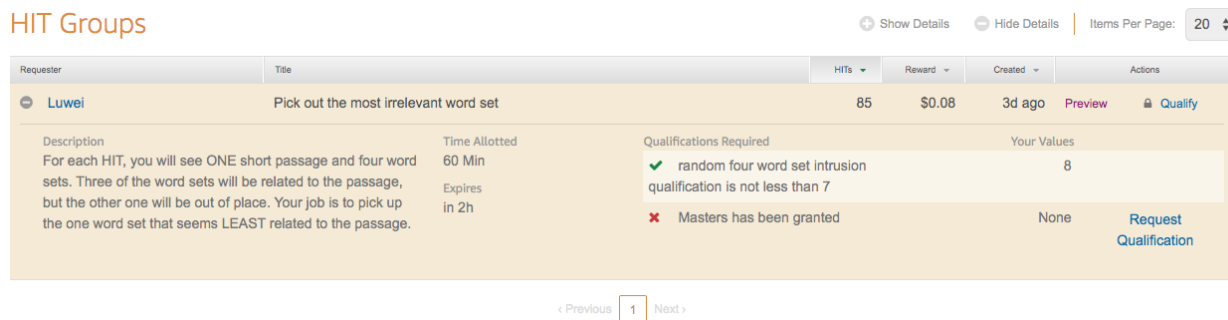
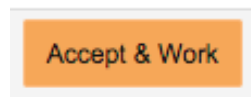


Figure SI20: How tasks look on Mturk



Once the workers get the qualification, the “lock” icon will become “Accept & work”.

Figure SI21: Workers can start working after seeing this “Accept & work” icon



Before they accept the HIT, they would **NOT** be able to see the actual options or document, which feature prevents them from selecting the tasks.

They’ll only be able to see the actual tasks after accepting. Notice that they can ALWAYS see the instructions on the left.

Researchers are able to extend the time of their HITs before or after the HITs expire.

```
for(i in HITids[[1]]){
  ExtendHIT(hit = i, add.seconds = 3600)
}
```

Figure SI22: Before Accepting the HIT

You must accept this Requester's HIT before working on it. [Learn more](#)

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. We understand that in many cases it will be hard to tell what the "right" choice is. Use your best judgement, but please be attentive. **We have randomly mixed in a number of gold-standard HITs whose answers are less ambiguous to evaluate the quality of your work.** If you miss a large number of gold-standard HITs, you may be **blocked** from continued participation in this study (and future studies).

After you've done a relatively large number of HITs of a specific task structure, your qualification might be temporarily suppressed. This does not mean you've done anything wrong, but we need to ensure that a variety of workers complete our tasks. This is **NOT a block** and will have **NO negative**

Please read the paragraph below and answer the following question

Please accept the HIT to see the actual paragraph

After reading the above passage, please click on the set of words below that is most unrelated to passage.

Option 1
 Option 2
 Option 3
 Option 4

Report this HIT ▾ | Why Report ▾

Skip Accept

Figure SI23: After Accepting the HIT

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. We understand that in many cases it will be hard to tell what the "right" choice is. Use your best judgement, but please be attentive. **We have randomly mixed in a number of gold-standard HITs whose answers are less ambiguous to evaluate the quality of your work.** If you miss a large number of gold-standard HITs, you may be **blocked** from continued participation in this study (and future studies).

After you've done a relatively large number of HITs of a specific task structure, your qualification might be temporarily suppressed. This does not mean you've done anything wrong, but we need to ensure that a variety of workers complete our tasks. This is **NOT a block** and will have **NO negative impact** on you Mturk record.

Instructions

For each HIT, you will see **one short**

Please read the paragraph below and answer the following question

Thank you to Missouri State University College of Agriculture for hosting the 2017 Agriculture Forum. Ill keep working to make sure our current and future Missouri farmers have the tools they need to take advantage of the great economic opportunities ahead.

After reading the above passage, please click on the set of words below that is most unrelated to passage.

forward, together, important, challengeand
 federal, congress, government, proposal
 working, across, week, look
 country, make, continue, surest

Submit

Report this HIT ▾ | Why Report ▾

Return

5.6 Retrieving and Evaluating Results

The `getResults()` function allows the researchers to retrieve results from a batch. It would work **no matter the batch has been completed or not**. In that sense, this function can also be used to check batch status. Notice that `batch_id` here can be any string that helps the researchers to refer to the batch in the future. It is first specified here. The `retry`, if `TRUE`, retries retrieving results from Mturk API at most five times, given that it fails retrieving all results the first time. The default is `retry = TRUE`.

```
testresults <- getResults(batch_id = "testbatch",
                          hit_ids = HITids,
                          retry = FALSE)
```

```
# Console returns
> Start getting HITs...
> 208 / 250 results retrieved
```

We highly recommend users to record the results immediately, especially when the batch is finished as the results might be deleted by the Mturk platform at some point.

To evaluate results, use `evalResults()`. This function identifies workers who consistently give poor-quality work and also returns the rate that human workers agree with the machine prediction.

```
evalResults(results = testresults,
            key = record,
            type = "T8WSI")
```

Figure SI24: Evaluating Results: Output from `evalResults()`

```
208 / 250 results will be evaluated
35 / 35 gold-standard HITs are answered correct
123 / 173 non-gold-standard HITs are answered correct
$`Gold-standard HIT Correct Rate`
[1] "1"          "35 / 35"

$`Gold-standard HIT Correct Rate by Workers`

      TRUE
0      0
A1 0.75 19
A2 0.5 8
A3 0.5 8

$`Gold-standard HIT Correct Rate`
[1] "0.710982658959538" "123 / 173"
```

In rare cases, researchers will encounter workers who keep giving work of poor quality. Therefore, researcher can reject HITs from these workers and ban them from participating in future tasks in two steps.

- 1) Identify these workers and revoke their qualifications.

```
workers_to_ban <- c("WORKER1", "WORKER2")
for(worker in workers_to_ban){
  AssignQualifications(qual = "QUALIFICATION_ID",
                       workers = worker,
                       notify = TRUE,
```

```
        value = 0)  
}
```

2) Reject their qualifications.

```
AssignmentsToReject <-  
  testresult$assignment_id[testresults$worker_id %in% workers_to_ban]  
RejectAssignments(assignments = unique(AssignmentsToReject),  
  feedback = "The quality of your work in the most recent  
  batch of HITs did not pass an audit. Your HITs from this  
  batch have been rejected and your qualification has  
  been deactivated.")
```

6 Workers' Training and Performance

This section first presents the training modules for the five tasks used in the paper. The second part provides some statistics and discussion about the workers' performance.

6.1 Training Module for Word Intrusion

Completing this training module qualifies you to complete Word Intrusion HITs.

Basic instructions

1. For each HIT, you will see FIVE words.
2. Four of them will be related to each other, but one word will be out of place.
3. Your job is to pick up the one word that does NOT belong with the others.

Background

The words you see are taken from U.S. senators' official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. It is critical that you read each word grouping carefully. Scanning or reading quickly will result in low-quality evaluations, and you may be blocked from continued participation in this study (and future studies).

Part 1

The next 5 questions are example HITs.

We will provide you with the correct answer. These will not be scored and will not count for or against your qualification.

Practice HIT #1

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- job
- business
- day
- work
- economy

Answer: The correct answer is “day”. “job”, “business”, “work” and “economy” are all related to economic activities. “day” is not.

Practice HIT #2

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- energy

- o water
- o climate
- o small
- o clean

Answer: The correct answer is “small”. “energy”, “water”, “climate” and “clean” are all connected to one another under the theme of environmental policy concerns. However, “small” is irrelevant to this theme.

Practice HIT #3

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o country
- o america
- o oil
- o nation
- o freedom

Answer: The correct answer is “oil”. “country”, “america”, “nation” and “freedom” are related to America and American values. However, “oil” is irrelevant to this theme.

Practice HIT #4

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o farm
- o school
- o students
- o university
- o high

Answer: The correct answer is “farm”. “school”, “students”, “university” and “high” are all related to education. However, “farm” is not.

Practice HIT #5

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o veterans
- o service
- o proud
- o meeting
- o force

Answer: The correct answer is “meeting”. “veterans”, “service”, “proud” and “force” are all related to veterans affairs and the armed services. However, “meeting” is irrelevant.

Part 2

The next 8 questions are your test HITs.

You must answer at least 7 of the test HITS correctly to receive the qualification.

Test HIT #1

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- tax
- budget
- debt
- bipartisan
- spending

Answer: [Not Available in the Real Test] “bipartisan”

Test HIT #2

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- homework
- security
- nuclear
- iran
- hearing

Answer: [Not Available in the Real Test] “homework”

Test HIT #3

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- watch
- news
- committee
- live
- show

Answer: [Not Available in the Real Test] “committee”

Test HIT #4

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- vegetable
- congress
- president
- republicans
- house

Answer: [Not Available in the Real Test] “vegetable”

Test HIT #5

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o health
- o bill
- o act
- o airplane
- o legislation

Answer: [Not Available in the Real Test] “airplane”

Test HIT #6

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o agriculture
- o happy
- o farmers
- o industry
- o trade

Answer: [Not Available in the Real Test] “happy”

Test HIT #7

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o sanders
- o social
- o night
- o billion
- o spending

Answer: [Not Available in the Real Test] “night”

Test HIT #8

Please read the five words below, and choose one that is most IRRELEVANT to the other four.

- o nuclear-power
- o discuss
- o watch
- o spoke
- o read

Answer: [Not Available in the Real Test] “nuclear-power”

Before you submit your answers...

You will only have 1 chance to take this test. Make sure that you are satisfied with all of your answers above before submitting.

If you become qualified to participate in the HITs, please continue to fully read each future HIT and provide your best guess of the correct answer. Your performance will be monitored as you complete more HITs. If you provide poor quality answers, you may be blocked from continued participation in this study (and future studies).

6.2 Training Module for Top 8 Word Set Intrusion

Completing this training module qualifies you to complete Top Eight Word Set Intrusion HITs.

Basic instructions

1. For each HIT, you will see ONE SHORT PASSAGE and FOUR word sets.
2. Three of the word sets will be related to the passage, but the other one will be out of place.
3. Your job is to pick up the one word set that seems LEAST related to the passage.

Background

The passages and words you see are taken from U.S. senators' official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. It is critical that you read each word grouping carefully. Scanning or reading quickly will result in low-quality evaluations, and you may be blocked from continued participation in this study (and future studies).

Part 1

The next 5 questions are example HITs.

We will highlight the important text and provide you with the correct answer. These will not be scored and will not count for or against your qualification.

Practice HIT #1

Heartening story in the Lawrence Journal-World about 89-year-old WWII Veteran and The University of Kansas Football Team Alumnus Bryan Sperry. He stole the show in Saturday's alumni flag football game.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o veterans, service, national, honor, military, thank, proud, women
- o great, students, school, office, visit, thanks, state, county
- o watch, senate, morning, last, discuss, live, read, news
- o sanders, tax, budget, new, bernie, said, debt, pay

Answer: The correct answer is “sanders, tax, budget, new, bernie, said, debt, pay”. The passage is about Bryan Sperry. It is a local interest story that focuses on the achievements of a veteran a football alumni involved in a flag football game.

The word set “veterans, service, national, honor, military, thank, proud, women” clearly relates to his identity as “an 89-year-old WWII Veteran.

The word set “great, students, school, office, visit, thanks, state, county” relates to his identity as an alumnus of the University of Kansas Football Team and the fact that this post is celebrating a local event/achievement.

The passage also references local news coverage (the post is actually about a news story), which is somewhat related to the set “watch, senate, morning, last, discuss, live, read, news.”

The word set “sanders, tax, budget, new, bernie, said, debt, pay”, however, seems to relate public policy and are not related to a local non-political event.

Practice HIT #2

Earlier this week I spoke at a town hall in support of Referendum 74 at First Baptist Church in Seattle. Stand with me – and thousands of Washington families – and click below to volunteer in support of marriage equality!

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o watch, senate, morning, last, discuss, live, read, news
- o day, family, people, one, life, every, years, world
- o energy, water, climate, change, oil, federal, clean, regulations
- o great, students, school, office, visit, thanks, state, county

Answer: The correct answer is “energy, water, climate, change, oil, federal, clean, regulations”. The passage is in support of a state referendum on gay marriage. It also references a town hall (a local political meeting) held at a church.

Two word sets “great, students, school, office, visit, thanks, state, county” and “watch, senate, morning, last, discuss, live, read, news” relates to the fact that the post references the Senator appearing at a local political gathering.

Another word set “day, family, people, one, life, every, years, world” seems to be related to the fact that the issue concerns “marriage” and “families.”

The final word set “energy, water, climate, change, oil, federal, clean, regulations”, however, seems to primarily be about environmental policy.

Practice HIT #3

The London Metropolitan Police Department has great taste in cars! They're patrolling the streets of London in BMW X5s made in – you guessed it – South Carolina.

I'm in London this weekend attending the annual Farnborough International Airshow in support of the Boeing 787 Dreamliners also made in South Carolina. #PalmettoPride #PalmettoPrideintheUK

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o new, jobs, state, help, businesses, work, business, economy
- o veterans, service, national, honor, military, thank, proud, women
- o president, congress, people, senate, american, americans, government, obama
- o great, students, school, office, visit, thanks, state, county

Answer: The correct answer is “president, congress, people, senate, american, americans, government, obama”. The passage expresses pride in the cars and airplanes made in South

Carolina. The passage mentions cars, the police, and Boeing airplanes manufactured in South Carolina, and an airshow in London.

One word set “new, jobs, state, help, businesses, work, business, economy” clearly relates to the economy, and is related to the cars and planes manufactured in South Carolina.

A second word set “great, students, school, office, visit, thanks, state, county” again relates to the local nature and focus of the post.

A third word set “veterans, service, national, honor, military, thank, proud, women” is a bit more difficult. Probably it is related to both the reference to the police and the international airshow.

The fourth set of words “president, congress, people, senate, american, americans, government, obama”, has no real connection with the passage.

Practice HIT #4

*Today’s **executive order** from **President Trump** is more about extreme **xenophobia** than extreme vetting. This **executive order** is the equivalent of a “Keep Out” sign posted at **America’s** borders.*

*Turning away **immigrants** based on their **nationality and religion** is **un-American** and in direct opposition to everything for which our Founding Fathers fought. **President Trump** may not call it a Muslim ban, but it is, and runs afoul of our morals and values.*

*We must ensure that we have the strongest safeguards in place to keep terrorists from ever reaching our shores. And we must fully and thoroughly vet all **refugees** to screen out any potential terrorist threats.*

*But as conflict and war force millions around the **world** from their homeland, the **United States** should welcome more refugees, not less. Suspending the **U.S.** refugee resettlement program will endanger refugees’ lives and tear **families** apart.*

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o day, family, people, one, life, every, years, world
- o great, students, school, office, visit, thanks, state, county
- o president, congress, people, senate, american, americans, government, obama
- o committee, security, hearing, secretary, united, states, department, senate

Answer: The correct answer is “great, students, school, office, visit, thanks, state, county”. The passage is about President Trump’s executive order on immigrants, which is a national-level policy discussion.

The word set “president, congress, people, senate, american, americans, government, obama” is about national politics and, therefore, relevant.

Another word set “committee, security, hearing, secretary, united, states, department, senate” seems to relate to politics and perhaps national security.

The word set “day, family, people, one, life, every, years, world” is a bit tricky. However, it seems to relate to social issues and is triggered by the mention of families.

The final word set “great, students, school, office, visit, thanks, state, county” seems to refer to local politics and local events and is not clearly related to the discussion of national

immigration policy.

Practice HIT #5

*Women in New Mexico and across the country deserve to make their own decision about **family planning**. **Laws** to protect a **women's right** to choose what's best for her **body** and **well-being** should not be restricted to what **state** she lives in. On the 42nd anniversary of the landmark *Roe v. Wade* decision, I remain committed to strengthening **women's reproductive rights** and freedom to make their own personal **health care** choices. #Roe42*

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o president, congress, people, senate, american, americans, government, obama
- o energy, water, climate, change, oil, federal, clean, regulations
- o health, bill, act, care, legislation, help, need, bipartisan
- o day, family, people, one, life, every, years, world

Answer: The correct answer is “energy, water, climate, change, oil, federal, clean, regulations”. The passage advocates for a pro-choice position and frames abortion as personal health care choices (family planning).

The word set “health, bill, act, care, legislation, help, need, bipartisan” is about health care and, therefore, relevant.

Another word set “day, family, people, one, life, every, years, world” seems to relate to social issues and is triggered by the mention of families.

The word set “President, congress, people, senate, american, americans, government, obama” is a bit more ambiguous. It relates to legislative politics and the passage is talking about an important issue in legislative politics. Thus, it is also relevant.

The final word set “energy, water, climate, change, oil, federal, clean, regulations” is about environmental issues and not relevant to the passage.

Part 2

The next 8 questions are your test HITs.

You must answer at least 7 of the test HITs correctly to receive the qualification.

Test HIT #1

Being conservative means controlling spending and costs. Those without insurance are forced to seek care in expensive ER's after their conditions have worsened. Hospitals are legally required to treat patients in ERs regardless of whether they can pay and they pass the cost of treatment onto the privately insured, increasing total costs. Increasing coverage and allowing more people to manage their health care is cheaper for society. The Cassidy-Collins would accomplish this.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o conservative, congress, people, senate, american, society, government, obama
- o spending, tax, budget, new, cost, price, debt, pay
- o military, veterans, service, national, defence, threat, attack, iran
- o health, bill, act, care, legislation, help, need, insurance

Answer: [Not Available in the Real Test] “military, veterans, service, national, defence, threat, attack, iran”

Test HIT #2

Today Sen. Judd Gregg and I introduced The Bipartisan Tax Fairness and Simplification Act of 2010, which will help middle-class taxpayers by streamlining and modernizing the outdated tax code. The proposal includes fiscally responsible tax cuts to help working families struggling to make ends meet and also eliminates the corporate tax break that encourages companies to invest overseas rather than creating jobs in the U.S.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o water, river, environment, change, new, future, proud, great
- o senate, bill, act, introduce, legislation, help, need, bipartisan
- o invest, jobs, state, help, company, work, business, economy
- o sanders, tax, budget, cut, bernie, said, debt, pay

Answer: [Not Available in the Real Test] “water, river, environment, change, new, future, proud, great”

Test HIT #3

Obama simultaneously could ruin Putin’s day and brighten the lives of millions of Americans. All Obama needs is the courage to tell the environmental Left to let him do the right thing.

Today is the last day for the public to comment on whether the State Department should approve a presidential permit for the #KeystoneXL pipeline. This article should give you some ideas of things to include in a comment. Here is where you can leave your comment: [link omitted]

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o day, family, people, public, life, every, million, world
- o energy, water, climate, change, oil, federal, clean, regulations
- o president, congress, people, senate, american, state, government, obama
- o veterans, service, national, honor, military, thank, proud, women

Answer: [Not Available in the Real Test] “veterans, service, national, honor, military, thank, proud, women”

Test HIT #4

Look, here is the bottom line. We remain the only nation in the industrialized world that doesn't guarantee health care to all people. We have 29 million people who are uninsured, yet we are spending far, far more per capita on health care than do the people of any other country.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o health, bill, act, care, legislation, help, need, bipartisan
- o students, school, office, class, young, educare, college, county
- o president, congress, people, senate, american, americans, government, obama
- o spending, tax, budget, new, bernie, said, debt, pay

Answer: [Not Available in the Real Test] “students, school, office, class, young, educare, college, county”

Test HIT #5

Like so many Rhode Islanders, I am deeply disturbed by President-elect Trump's appointment of Steve Bannon to be his chief strategist in the White House. As Executive Chairman of Breitbart, Bannon served as a conduit for some of the worst sentiments in our society – hatred and violence on the basis of race, religion, gender, and way of life. As CEO of Trump's campaign, he employed the hateful code of the white supremacist movement to leverage prejudice and fear, for political gain. Now, the President-elect wants Bannon to help guide his administration. Many Rhode Islanders have called and written to tell me how unsettling Bannon's presence in the White House is, and I share their concern. If President-elect Trump will not denounce the dangerous ideas Steve Bannon has represented, it is up to the rest of us to stand firm against them.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o students, school, office, class, young, educare, college, county
- o president, congress, people, senate, american, americans, government, obama
- o day, family, people, one, life, every, equal, world
- o committee, appointment, hearing, represent, united, states, department, senate

Answer: [Not Available in the Real Test] “students, school, office, class, young, educare, college, county”

Test HIT #6

I wish Democrats would show some interest in how and why President Obama and Susan Rice got it so wrong about the true nature of the Benghazi attack. I wish Democrats would show a little interest about why Secretary Clinton was clueless about the multiple security requests coming from Benghazi and how she allowed our mission to become a death trap. I wish Democrats would show a little interest in finding out whether Mike Morell, the former #2 at the CIA, lied to Congress and the American people about a protest that never happened. Democrats seem

to be more interested in protecting the Obama Administration than they are in getting the truth.

If Republicans win a Senate majority in 2014, one of the first things I will insist on is hearings that actually get to the bottom of what happened before, during, and after the Benghazi attack.

The families of those who lost loved ones, and the American people, deserve nothing less than a full accounting.

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o day, family, people, one, life, every, years, world
- o committee, security, hearing, secretary, united, states, department, senate
- o president, congress, people, senate, american, americans, government, obama
- o company, jobs, invest, help, business, work, unemployment, economy

Answer: [Not Available in the Real Test] “company, jobs, invest, help, business, work, unemployment, economy”

Test HIT #7

Tomorrow, I'll be attending the White House bipartisan summit on health care reform hosted by President Obama. Share your ideas on health care reform with me now – so I can share your feedback with President Obama and Congressional leaders on Thursday!

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o military, veterans, service, national, defence, threat, attack, iran
- o health, bill, act, care, legislation, help, need, bipartisan
- o president, congress, people, senate, american, americans, government, obama
- o watch, senate, morning, share, discuss, live, feedback, news

Answer: [Not Available in the Real Test] “military, veterans, service, national, defence, threat, attack, iran”

Test HIT #8

Today Chairman Johnson held a hearing on duplication, waste, and fraud in federal programs.

Seven years later, with less than half of GAO's recommendations even implemented, GAO estimates that this report has resulted in actual savings of \$75 billion. One simple idea has saved American taxpayers tens of billions of dollars. However, there are still hundreds of recommendations that have gone unimplemented, and very little actual duplication in the federal government has been addressed. I am pleased that the Trump administration is taking this problem seriously. The Executive Order signed by President Trump and the memorandum by OMB Director Mulvaney that followed will result in a plan to reorganize and

*streamline the federal government and help it better serve the American people.
This is long overdue.*

After reading the above passage, please click on the set of words below that is most UNRELATED to passage.

- o sanders, tax, budget, new, bernie, said, debt, pay
- o committee, order, hearing, federal, united, states, department, senate
- o president, administration, congress, people, senate, americans, government, obama
- o students, school, office, class, young, educate, college, county

Answer: [Not Available in the Real Test] “students, school, office, class, young, educate, college, county”

Before you submit your answers...

You will only have 1 chance to take this test. Make sure that you are satisfied with all of your answers above before submitting.

If you become qualified to participate in the HITs, please continue to fully read each future HIT and provide your best guess of the correct answer. Your performance will be monitored as you complete more HITs. If you provide poor quality answers, you may be blocked from continued participation in this study (and future studies).

6.3 Training Module for Random 4 Word Set Intrusion

Completing this training module qualifies you to complete Random Four Word Set Intrusion HITs

Basic instructions

1. For each HIT, you will see FOUR word sets.
2. Three of the word sets will be related to one another, but the other one will be out of place.
3. Your job is to pick up the one word set that seems LEAST related to others.

Background

The words you see are taken from U.S. senators’ official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. It is critical that you read each word grouping carefully. Scanning or reading quickly will result in low-quality evaluations, and you may be blocked from continued participation in this study (and future studies).

Part 1

The next 5 questions are example HITs.

We will provide you with the correct answer. These will not be scored and will not count for or against your qualification.

Practice HIT #1

Please click on the word set below that is most UNRELATED to the other three.

- o bill, govern, vote, cut
- o senate, congress, pass, plan
- o tax, budget, house, reform
- o report, today, press, emergency

Answer: The correct answer is “respond, today, first-aid, emergency”. While three word sets are related to policy-making institutions and procedures (“bill, govern, vote, cut”, “senate, congress, pass, plan”, “tax, budget, house, reform”), the set of words “respond, today, first-aid, emergency” is not. Rather it seems to relate to time sensitive response and emergencies.

Practice HIT #2

Please click on the word set below that is most UNRELATED to the other three.

- o great, thank, visit, city
- o state, enjoy, tour, center
- o reform, obamacare, federal, year
- o happy, team, celebrate, park

Answer: The correct answer is “reform, obamacare, federal, year”. While the other three word sets relate to celebration of local events (“great, thank, visit, city”, “state, enjoy, tour, center”, and “happy, team, celebrate, park”), the set of words “reform, obamacare, federal, year” is not but rather relates to health care reform policy.

Practice HIT #3

Please click on the word set below that is most UNRELATED to the other three.

- o nation, student, service, school
- o honor, education, veteran, state
- o social, call, secure, post
- o learn, high, college, proud

Answer: The correct answer is “social, call, secure, post”. Three of the word sets (“nation, student, service, school”, “honor, education, professor, state”, and “learn, high, college, proud”) relate to education. The other word set (“social, call, secure, post”) is not.

Practice HIT #4

Please click on the word set below that is most UNRELATED to the other three.

- o day, family, country, people
- o proud, learn, university, service
- o women, american, live, make
- o work, life, nation, love

Answer: The correct answer is “proud, learn, university, college”. While the three other word sets seems to be generally about American ideals and values (“day, family, country, people”, “women, american, live, make”, and “work, life, nation, love”), the set of words

“proud, learn, university, college” is instead more closely related to education.

Practice HIT #5

Please click on the word set below that is most UNRELATED to the other three.

- o job, business, help, work
- o confirm, judge, nominate, rule
- o energy, economy, state, new
- o develop, continue, import, protect

Answer: The correct answer is “confirm, judge, nominate, rule”. While the other three word sets are related to the economy and businesses (“job, business, help, work”, “energy, economy, state, new” and “develop, continue, import, protect”), the set of words “confirm, judge, nominate, rule” is more clearly related to judicial appointment and legal issues.

Part 2

The next 8 questions are your test HITS.

You must answer at least 7 of the test HITS correctly to receive the qualification.

Test HIT #1

Please click on the word set below that is most UNRELATED to the other three.

- o discuss, join, week, senate
- o hear, talk, share, live
- o many, world, nation, america
- o office, morning, watch, meet

Answer: [Not Available in the Real Test] “many, world, nation, america”

Test HIT #2

Please click on the word set below that is most UNRELATED to the other three.

- o thank, visit, good, time
- o health, care, veteran, help
- o need, provide, support, work
- o drug, program, prevent, ensure

Answer: [Not Available in the Real Test] “thank, visit, good, time”

Test HIT #3

Please click on the word set below that is most UNRELATED to the other three.

- o right, law, court, senate
- o vote, justice, supreme, constitution
- o judge, nominate, rule, investigate
- o news, report, read, said

Answer: [Not Available in the Real Test] “news, report, read, said”

Test HIT #4

Please click on the word set below that is most UNRELATED to the other three.

- o president, action, security, administration
- o elect, decision, political, confirm
- o nation, threat, trump, deal
- o secretary, congress, immigrant, defense

Answer: [Not Available in the Real Test] “elect, decision, political, confirm”

Test HIT #5

Please click on the word set below that is most UNRELATED to the other three.

- o family, legislation, children, bill
- o news, today, said, report
- o governor, state, press, call
- o federal, emergency, post, continue

Answer: [Not Available in the Real Test] “family, legislation, children, bill”

Test HIT #6

Please click on the word set below that is most UNRELATED to the other three.

- o west, proud, community, learn
- o university, force, military, nation
- o student, service, member, thank
- o veteran, help, need, support

Answer: [Not Available in the Real Test] “veteran, help, need, support”

Test HIT #7

Please click on the word set below that is most UNRELATED to the other three.

- o administration, deal, threat, state
- o small, support, community, economy
- o create, continue, import, protect
- o job, business, help, work

Answer: [Not Available in the Real Test] “administration, deal, threat, state”

Test HIT #8

Please click on the word set below that is most UNRELATED to the other three.

- o justice, elect, supreme, political
- o decision, protect, general, constitution
- o right, law, court, senate
- o bill, american, tax, people

Answer: [Not Available in the Real Test] “bill, american, tax, people”

Before you submit your answers...

You will only have 1 chance to take this test. Make sure that you are satisfied with all of your answers above before submitting.

If you become qualified to participate in the HITs, please continue to fully read each future HIT and provide your best guess of the correct answer. Your performance will be monitored as you complete more HITs. If you provide poor quality answers, you may be blocked from continued participation in this study (and future studies).

6.4 Training Module for Label Intrusion

Completing this training module qualifies you to complete Label Intrusion HITs

Basic instructions

1. For each HIT, you will see ONE SHORT PASSAGE and FOUR labels.
2. Three of the labels will be related to the passage, but the other one will be out of place.
3. Your job is to pick up the one label that seems LEAST related to the passage.

Background

The passages you see are taken from U.S. senators' official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

Using a computer algorithm, we divided these posts into different categories based on the words they contain. The labels you see are our tentative summaries of those issues or topics.

We want to see if you can identify the label that does not belong with the post.

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. It is critical that you read each word grouping carefully. Scanning or reading quickly will result in low-quality evaluations, and you may be blocked from continued participation in this study (and future studies).

Part 1

The next 5 questions are example HITs.

We will highlight the important text and provide you with the correct answer. These will not be scored and will not count for or against your qualification.

Practice HIT #1

*Alexander Cosponsor's **Cut, Cap, and Balance Act**: At a time when we're borrowing 40 cents of every dollar we spend, I will support serious **proposals** like the "Cut, Cap and Balance Act" to reduce out-of-control **Washington spending**. The final version of any such **legislation** should have an appropriate balance between reductions in both entitlement and discretionary spending, but this **bill** is a good start.*

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Government Spending
- o Legislation
- o American Politics
- o Emergency

Answer: The correct answer is "Emergency". The passage talks about a legislation on the federal budget cut. Thus, the labels "Legislation" and "Government Spending" are clearly

relevant. This is an important issue in “American Politics,” which is also relevant as a label. The remaining label “Emergency,” however, is out of place because the budget cut is usually not an emergent event.

Practice HIT #2

*Earlier this week I spoke at a **town hall** in support of **Referendum 74** at First Baptist Church in Seattle. Stand with me – and thousands of Washington **families** – and click below to volunteer in support of **marriage equality!***

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Local Events
- o Public Meeting
- o Environment
- o Human Wellbeing

Answer: The correct answer is “Environment”. The passage is in support of a state referendum on gay marriage. It also references a town hall (a local political meeting) held at a church. Two word sets “Local Events” and “Public meeting” relate to the fact that the post references the Senator appearing at a local political gathering. Another word set “Human Wellbeing” seems to be related to the fact that the issue concerns “marriage” and “families.” The final label “Environment,” however, is not mentioned.

Practice HIT #3

*The London Metropolitan **Police Department** has great taste in **cars!** They’re **patrolling** the streets of London in BMW X5s made in – you guessed it – **South Carolina**.*

*I’m in London this weekend attending the annual Farnborough International **Airshow** in support of the **Boeing 787 Dreamliners** also made in **South Carolina**.
#PalmettoPride #PalmettoPrideintheUK*

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Economy
- o Local Events
- o Presidential Politics
- o Police

Answer: The correct answer is “Presidential Politics”. The passage expresses pride in the cars and airplanes made in South Carolina. Those products are about “Economy.” South Carolina is local and the airshow belongs to “Local Events.” Since the police department is mentioned, the label “Police” should also be relevant. “Presidential Politics” has no real connection with the passage.

Practice HIT #4

Today's *executive order* from *President Trump* is more about extreme *xenophobia* than extreme vetting. This *executive order* is the equivalent of a "Keep Out" sign posted at *America's* borders.

Turning away *immigrants* based on their *nationality and religion* is *un-American* and in direct opposition to everything for which our *Founding Fathers* fought. *President Trump* may not call it a Muslim ban, but it is, and runs afoul of our morals and values.

We must ensure that we have the strongest safeguards in place to keep terrorists from ever reaching our shores. And we must fully and thoroughly vet all *refugees* to screen out any potential terrorist threats.

But as conflict and war force millions around the *world* from their homeland, the *United States* should welcome more *refugees*, not less. Suspending the *U.S. refugee* resettlement program will endanger refugees' lives and tear *families* apart.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Immigration
- o Local Events
- o Presidential Politics
- o Security and Defence

Answer: The correct answer is "Local Events". The passage is about President Trump's executive order on immigrants. The labels "Presidential Politics" and "Immigration" are thus clearly relevant. The sentence "We must ensure that we have the strongest safeguards in place to keep terrorists from ever reaching our shores." implies that "Security and Defence" are also concerned. "Local Events" is not clearly related to the discussion of national immigration policy.

Practice HIT #5

Women in New Mexico and across the country deserve to make their own decision about *family planning*. *Laws* to protect a *women's right* to choose what's best for her *body* and *well-being* should not be restricted to what *state* she lives in. On the 42nd anniversary of the landmark *Roe v. Wade* decision, I remain committed to strengthening *women's reproductive rights* and freedom to make their own personal *health care* choices. #Roe42

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Legislation
- o Environment
- o Healthcare
- o Daily Life

Answer: The correct answer is "Environment". The passage advocates for a pro-choice position in legislative politics and thus "Legislation" is clearly relevant. It frames abortion as personal "Healthcare" choices (family planning). The label "Daily Life" is triggered by

the mention of families. “Environment” is not relevant to the passage.

Part 2

The next 8 questions are your test HITs.

You must answer at least 7 of the test HITs correctly to receive the qualification.

Test HIT #1

The unexpected and tragic attacks on Pearl Harbor that occurred 75 years ago today—a date that even now still lives in infamy—triggered America to go to war. As a result of the unprovoked attacks in 1941, many of our brave servicemen and women lost their lives. Today and every year on December 7th, I hope we each take a moment to reflect on and pay tribute to the sacrifices made by those who were injured or lost at Pearl Harbor. And let us each re-dedicate ourselves to ensuring their lives were not lost in vain by honoring and upholding the quintessential American values and freedoms they were fighting for.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o American History
- o Security and Defence
- o Education
- o Honoring Veterans

Answer: [Not Available in the Real Test] “Education”

Test HIT #2

Today Sen. Judd Gregg and I introduced The Bipartisan Tax Fairness and Simplification Act of 2010, which will help middle-class taxpayers by streamlining and modernizing the outdated tax code. The proposal includes fiscally responsible tax cuts to help working families struggling to make ends meet and also eliminates the corporate tax break that encourages companies to invest overseas rather than creating jobs in the U.S.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Immigration
- o Legislation
- o Economy
- o Tax and Budget

Answer: [Not Available in the Real Test] “Immigration”

Test HIT #3

Obama simultaneously could ruin Putin's day and brighten the lives of millions of Americans. All Obama needs is the courage to tell the environmental Left to let him do the right thing.

Today is the last day for the public to comment on whether the State Department should approve a presidential permit for the #KeystoneXL pipeline. This article should give you some ideas of things to include in a comment. Here is where you can leave your comment: [link omitted]

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- Human Wellbeing
- Environment
- Presidential Politics
- Honoring Veterans

Answer: [Not Available in the Real Test] "Honoring Veterans"

Test HIT #4

Look, here is the bottom line. We remain the only nation in the industrialized world that doesn't guarantee health care to all people. We have 29 million people who are uninsured, yet we are spending far, far more per capita on health care than do the people of any other country.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- Healthcare
- Education
- Government Spending
- Human Wellbeing

Answer: [Not Available in the Real Test] "Education"

Test HIT #5

In Warroad, MN, Hockeytown USA, with 3 proud U.S. Hockey silver medalists: Rubin Bjorkman (1952), Gordon Christian (1956) and Henry Boucha (1972). Warroad boasts 7 Olympic hockey medalists with Gigi Marvin and TJ Oshie set to win more. Oshie's grandpa played on 1948 Warroad team! Go USA hockey!

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- Healthcare
- Announcement
- Game
- Celebration

Answer: [Not Available in the Real Test] "Healthcare"

Test HIT #6

I wish Democrats would show some interest in how and why President Obama and Susan Rice got it so wrong about the true nature of the Benghazi attack. I wish Democrats would show a little interest about why Secretary Clinton was clueless about the multiple security requests coming from Benghazi and how she allowed our mission to become a death trap. I wish Democrats would show a little interest in finding out whether Mike Morell, the former #2 at the CIA, lied to Congress and the American people about a protest that never happened. Democrats seem to be more interested in protecting the Obama Administration than they are in getting the truth.

If Republicans win a Senate majority in 2014, one of the first things I will insist on is hearings that actually get to the bottom of what happened before, during, and after the Benghazi attack.

The families of those who lost loved ones, and the American people, deserve nothing less than a full accounting.

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o The Public
- o Political Parties
- o Presidential Politics
- o Economy

Answer: [Not Available in the Real Test] “Economy”

Test HIT #7

Tomorrow, I'll be attending the White House bipartisan summit on health care reform hosted by President Obama. Share your ideas on health care reform with me now – so I can share your feedback with President Obama and Congressional leaders on Thursday!

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Honoring Veterans
- o Healthcare
- o American Politics
- o Public Participation

Answer: [Not Available in the Real Test] “Honoring Veterans”

Test HIT #8

Yesterday I met with Anjali Lall and Maneesh Apte, U.S. Presidential Scholar recipients who recently graduated from Davies High School in Fargo, ND. The United States Presidential Scholar Program annually awards up to 141 high school graduates nationwide, honoring students who demonstrate exceptional academic achievement. Congratulations Anjali and Maneesh!

Please read the four labels below and click on the label that is most UNRELATED to the passage.

- o Honorary Statement
- o Education
- o President
- o Economy

Answer: [Not Available in the Real Test] “Economy”

Before you submit your answers...

You will only have 1 chance to take this test. Make sure that you are satisfied with all of your answers above before submitting.

If you become qualified to participate in the HITs, please continue to fully read each future HIT and provide your best guess of the correct answer. Your performance will be monitored as you complete more HITs. If you provide poor quality answers, you may be blocked from continued participation in this study (and future studies).

6.5 Training Module for Optimal Label

Completing this training module qualifies you to complete Optimal Label HITs.

Basic instructions

1. For each HIT, you will see ONE SHORT PASSAGE and FOUR labels.
2. Your job is to pick up the one label that seems BEST summarizing the passage.

Background

The passages you see are taken from U.S. senators’ official Facebook postings. These postings can be on any issue or topic. Some postings are about certain legislation, some are about federal or statewide events, and some are about the administration, etc.

Using a computer algorithm, we divided these posts into different categories based on the words they contain. The labels you see are our tentative summaries of those issues or topics.

We want to see if you can identify the label that best summarizes the post.

Attention

Some of these choices will be very clear, but others will require you to use your best judgment. It is critical that you read each word grouping carefully. Scanning or reading quickly will result in low-quality evaluations, and you may be blocked from continued participation in this study (and future studies).

Part 1

The next 5 questions are example HITs.

We will highlight the important text and provide you with the correct answer. These will not be scored and will not count for or against your qualification.

Practice HIT #1

Alexander Cosponsor's *Cut, Cap, and Balance Act*: At a time when we're borrowing 40 cents of every dollar we spend, I will support serious *proposals* like the "Cut, Cap and Balance Act" to reduce out-of-control Washington spending. The final version of any such *legislation* should have an appropriate balance between reductions in both entitlement and discretionary spending, but this *bill* is a good start.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Insurance
- o Legal/Law
- o Immigration
- o Budget Cut

Answer: The correct answer is "Budget Cut". The passage talks about a legislation on the federal budget cut. The label "Budget Cut" best summarizes it. "Legal/Law", "Immigration", and "Immigration" are not relevant.

Practice HIT #2

In Warroad, MN, *Hockeytown USA*, with 3 proud U.S. Hockey silver *medalists*: Rubin Bjorkman (1952), Gordon Christian (1956) and Henry Boucha (1972). Warroad boasts 7 Olympic hockey medalists with Gigi Marvin and TJ Oshie set to win more. Oshie's grandpa played on 1948 Warroad *team*! *Go USA* hockey!

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Economy
- o Holiday Greetings
- o Celebration Messages
- o Information Sources

Answer: The correct answer is "Celebration Messages". The passage celebrates a town called "Warroad" for having many hockey medalists. "Celebration Messages" is thus the optimal label. The purpose of this message is not greeting people, so "Holiday Greetings" is not the best answer. While providing people with some information, the passage is not mainly about "Information Sources." "Economy" is not relevant.

Practice HIT #3

Yesterday I met with Anjali Lall and Maneesh Apte, U.S. Presidential *Scholar recipients* who recently *graduated* from Davies High *School* in Fargo, ND. The United States Presidential Scholar Program annually awards up to 141 *high school graduates* nationwide, *honoring students* who demonstrate exceptional *academic achievement*. Congratulations Anjali and Maneesh!

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Law Enforcement
- o Honoring Veterans
- o Education
- o Presidential Politics

Answer: The correct answer is “Education”. The passage congratulates a student for receiving the U.S. Presidential Scholar Award. It is clearly an educational issue and, thus, the label “Education” is optimal. Although the award’s name has the word, presidential, in it, the passage is not about presidential politics. “Law Enforcement” and “Honoring Veterans” are not implied.

Practice HIT #4

*The unexpected and tragic attacks on Pearl Harbor that occurred 75 years ago today—a date that even now still lives in infamy—triggered **America** to go to war. As a result of the unprovoked attacks in 1941, many of our brave **servicemen and women lost their lives**. Today and every year on December 7th, I hope we each take a moment to reflect on and **pay tribute** to the sacrifices made by those who were injured or lost at Pearl Harbor. And let us each re-dedicate ourselves to ensuring their lives were not lost in vain by honoring and upholding the quintessential **American values and freedoms** they were fighting for.*

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Terrorist Attack
- o Honoring Veterans
- o Water Pollution
- o American History

Answer: The correct answer is “Honoring Veterans”. The passage honors victims of the attacks on Pearl Harbor and specifically mentions “brave servicemen and women.” “Honoring Veterans” is the best summary. Although the attack took place 75 years ago, this message is not mainly about “American History.” The Pearl Harbor attack is not a “Terrorist Attack.” “Water Pollution” is not relevant.

Practice HIT #5

*The Red River Corridor Fund received more than \$1.1 million in federal funding to incentivize **private companies** to help **small businesses and local economies** grow. When our **small businesses** have an environment they can thrive in, it creates more **jobs** and helps **grow our communities and state**. These funds will help encourage more **private-public partnerships across the state** so **small businesses**, private lenders, and the federal government can work collaboratively to build North Dakota’s already growing **economy**.*

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Unemployment Rate
- o Economy
- o Local Issues
- o Healthcare

Answer: The correct answer is “Economy”. The passage advertises that North Dakota received federal funds supporting small businesses and local economies. “Economy” is relevant. “Unemployment Rate” is attempting given that the passage mentioned “creates more

jobs”. However, it is not optimal because it cannot cover other aspects of the economy such as “funding” and “private-public partnerships.” “Local Issues” is too vague. “Healthcare” is not relevant.

Part 2

The next 8 questions are your test HITs.

You must answer at least 7 of the test HITs correctly to receive the qualification.

Test HIT #1

I’m hosting a live telephone town hall call this Monday, April 24 at 5:45pm MT. If you’re available to talk on the phone, sign up to get a call. You can also participate and ask questions by streaming the event online. Click on the link below to do both: [link omitted]

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Drug Abuse
- o Legislative Politics
- o Education
- o Town Hall Meeting

Answer: [Not Available in the Real Test] “Town Hall Meeting”

Test HIT #2

Today, Congress has unanimously passed the Clay Hunt SAV Act to improve veteran suicide prevention programs. We must do everything we can to ensure our veterans and military families have access to high-quality mental health services. By improving treatment, intervention, and outreach, the Clay Hunt SAV Act will help more veterans get the help that they need and have earned. No veteran should ever have to wait for critical mental health care. Learn more: [link omitted]

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Veteran Healthcare
- o Economy
- o Drug Abuse
- o Nation Building

Answer: [Not Available in the Real Test] “Veteran Healthcare”

Test HIT #3

Today I met with Judge Neil Gorsuch, nominee to serve on the U.S. Supreme Court. Judge Gorsuch is a smart, diligent and thoughtful jurist. I am impressed by his remarkable commitment to the Constitution and the separation of power our founders envisioned. Oklahoma is within the jurisdiction of the 10th Circuit

Court of Appeals, on which Judge Gorsuch sits, and I have seen his judicial philosophy first hand. In Burwell v. Hobby Lobby, he wrote a concurring opinion in favor of the Oklahoma company's position, and upholding our First Amendment right to religious freedom. Judge Gorsuch is a qualified mainstream jurist who is well respected by conservatives and liberals alike and I urge Senate Democrats to allow and up or down vote on Gorsuch. I look forward to confirming him to the bench.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Supreme Court
- o Congratulation Messages
- o Military
- o Religious Freedom

Answer: [Not Available in the Real Test] "Supreme Court"

Test HIT #4

Recent actions by Iran have once again demonstrated their reluctance to curb their confrontational nuclear missile program. These dangerous, provocative actions by a country considered a state sponsor of terror should carry consequences. The decision to increase sanctions should come at no surprise. A line was drawn, and Iran deliberately defied the agreement. I stand by the President on his decision to implement a quick and deliberate strategy. It should send a clear message that Iran's reckless behavior will not be tolerated.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Justice
- o Security
- o Honoring Veterans
- o Government Spending

Answer: [Not Available in the Real Test] "Security"

Test HIT #5

Deadline extended: January 23 is the new deadline to apply for federal disaster assistance from Hurricane Matthew. Register with FEMA Federal Emergency Management Agency online: [link omitted]

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Emergency Management
- o Holiday Greetings
- o Foreign Policy
- o Public Hearing

Answer: [Not Available in the Real Test] "Emergency Management"

Test HIT #6

Interested in attending one of our nation’s military academies? My office is currently accepting applications from North Dakota students interested in attending the U.S. Military Academy at West Point, the U.S. Naval Academy, U.S. Air Force Academy or the U.S Merchant Marine Academy. The deadline is Oct. 5.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Healthcare
- o Doctor
- o American History
- o Application

Answer: [Not Available in the Real Test] “Application”

Test HIT #7

Today I brought together local business development and community leaders to discuss the essential role of rural communities in North Dakota’s economic success and ways to support new investment and growth in small towns across the state. Energy development in the Bakken has created unprecedented growth but we have to make sure this development’s success flows across the entire state. Rural communities like Ellendale are an important part of that effort, and by supporting hardworking North Dakotans and small businesses in these areas, we can continue to sustainably grow our state for future generations.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Local Businesses
- o Nation Building
- o Infrastructure Construction
- o Public Meeting

Answer: [Not Available in the Real Test] “Local Businesses”

Test HIT #8

While under oath during his confirmation hearing, Attorney General Jeff Sessions misled his colleagues in U.S. Senate about his communications with Russia Ambassador Sergey Kislyak. He should resign immediately. This is exactly why we need an independent special counsel to investigate the Russian hacking of our presidential election and any ties the Trump campaign may have had to Russia.

Please read the four labels below and click on the label that BEST summarizes the passage.

- o Family Memory
- o Information Sources
- o Games
- o Legal Institutions and Offices

Answer: [Not Available in the Real Test] “Legal Institutions and Offices”

Before you submit your answers...

You will only have 1 chance to take this test. Make sure that you are satisfied with all of your answers above before submitting.

If you become qualified to participate in the HITs, please continue to fully read each future HIT and provide your best guess of the correct answer. Your performance will be monitored as you complete more HITs. If you provide poor quality answers, you may be blocked from continued participation in this study (and future studies).

6.6 Workers’ Performance

To assess workers’ performance, we randomly mixed in a gold-standard HIT into every ten HITs. Once posted a batch, we checked its progress sporadically. We suppressed the qualification of workers who have missed more than 2 gold-standard HITs or who have done a relatively large number of HITs of a specific task structure. This operation has no negative impact on their Mturk records. Among all workers who have completed HITs in the batches demonstrated in the main manuscript, we have rejected and replaced work from only three of them who missed more than 4 gold-standard HITs each. These thresholds (2 and 4) are not iron rules. Future researchers can decide on their own standards based on the difficulty of their designed gold-standard HITs.

The gold-standard HIT approach yields consistent results. In the main manuscript, none of the identical pairs we presented are significantly different from each other.

Another way to assess the consistency of workers’ performance across identical trails is to assess the “agreement rate,” which we present in Tables SI6 and SI7. The numbers reflect the frequency that workers from the two trials get the task either right or wrong simultaneously, as opposed to one get it right but the other get it wrong.

Table SI6: Workers’ Agreement Rate in the Topic Validation Tasks

	Model 1	Model 2	Model 3	Model 4	Model 5
WI	0.816	0.696	0.724	0.696	0.684
T8WSI	0.62	0.606	0.68	0.72	0.704
R4WSI	0.714	0.754	0.85	0.884	0.728

We realize that these numbers are hard to interpret. In general, a high agreement rate is desirable. However, the agreement rate would, theoretically, be a function of the correct rate. This introduces a floor effect. Consider the case, for instance, where workers got 95% of the HITs correct, agreement rates can only go as low as 90%. For this reason, we did not include these tables in the main text, but only as supplementary information.

Table SI7: Workers’ Agreement Rate in the Label Validation Tasks

	Careful Coder	Cursory Coder
LI-Within	0.764	0.746
LI-Across	0.864	0.896
OL-Within	0.836	0.824
OL-Across	0.904	0.856

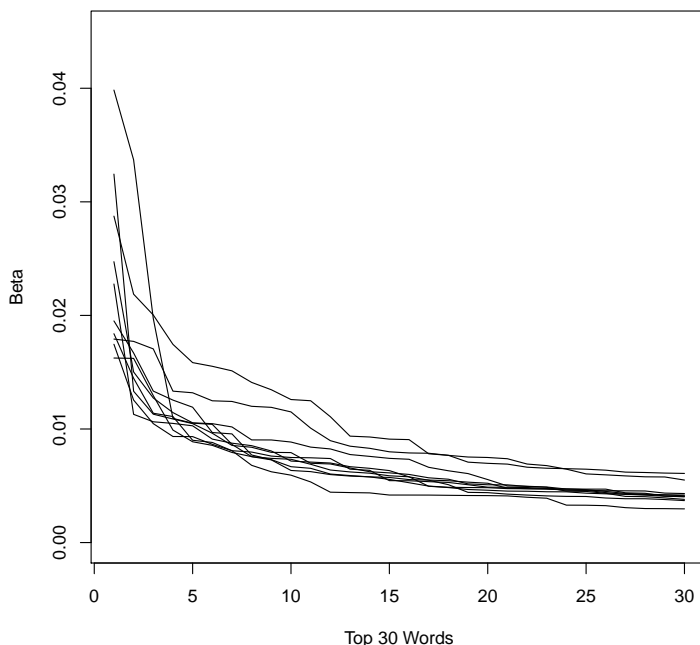
7 More on the Corpus, Topic Model Fit, and Labeling

This section presents more background information and discussions on the structural topic models in the paper and their labeling process. Specifically, we first discuss the distributions of word probability in the topic models, which is the reason we chose to draw words based on their probability in the topic validation tasks. We then present some word clouds to give people a different perspective of the topics in different topic models. Finally, we present the representative documents based on which we came up with the topic labels.

7.1 Word Mass Distribution

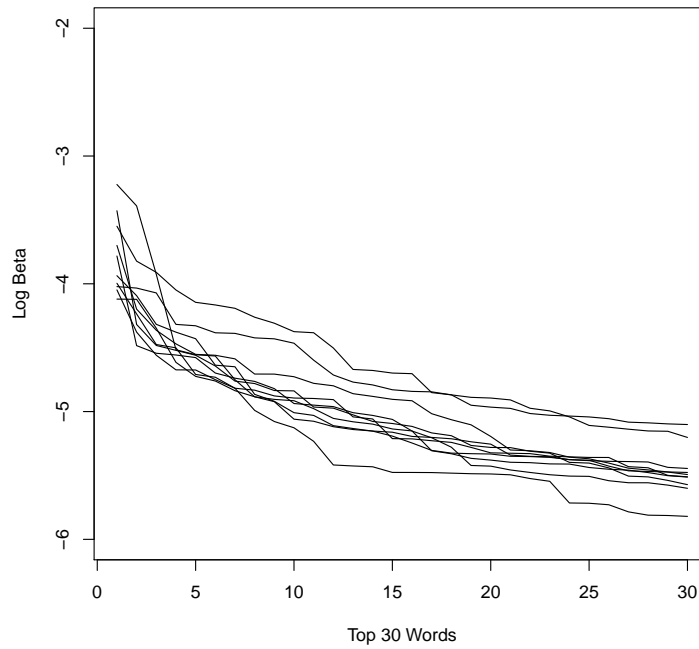
One major difference between our topic validation tasks and those from Chang et al. (2009) is that we randomly draw words based on their probabilities in a given topic from a given model. We made this choice because the word probabilities could vary a lot among the top 20 high probability words. Figures SI25 and SI26 demonstrate this point with Model 2 (the 10-topic model) from our paper, where `beta` refers to the word probability.

Figure SI25: Word Mass: Beta



Furthermore, Figure SI27 depicts the total word probability covered by the top n words for the four converged topic models (Model2, Model3, Model4, and Model5) from our paper. The shapes of the curves are similar, although the specific numbers vary across different models. On average, except Model2, the top 20 words cover about 50% of the total word probability.

Figure SI26: Word Mass: Log-Beta



7.2 Word Clouds

For the four structural topic models validated in the paper, we randomly select 6 topics from each of them and present the word clouds below. Just from the word cloud, it does not seem obvious that the four converged models (Model2, Model3, Model4, and Model5) are of different quality.

Figure SI27: Word versus Mass

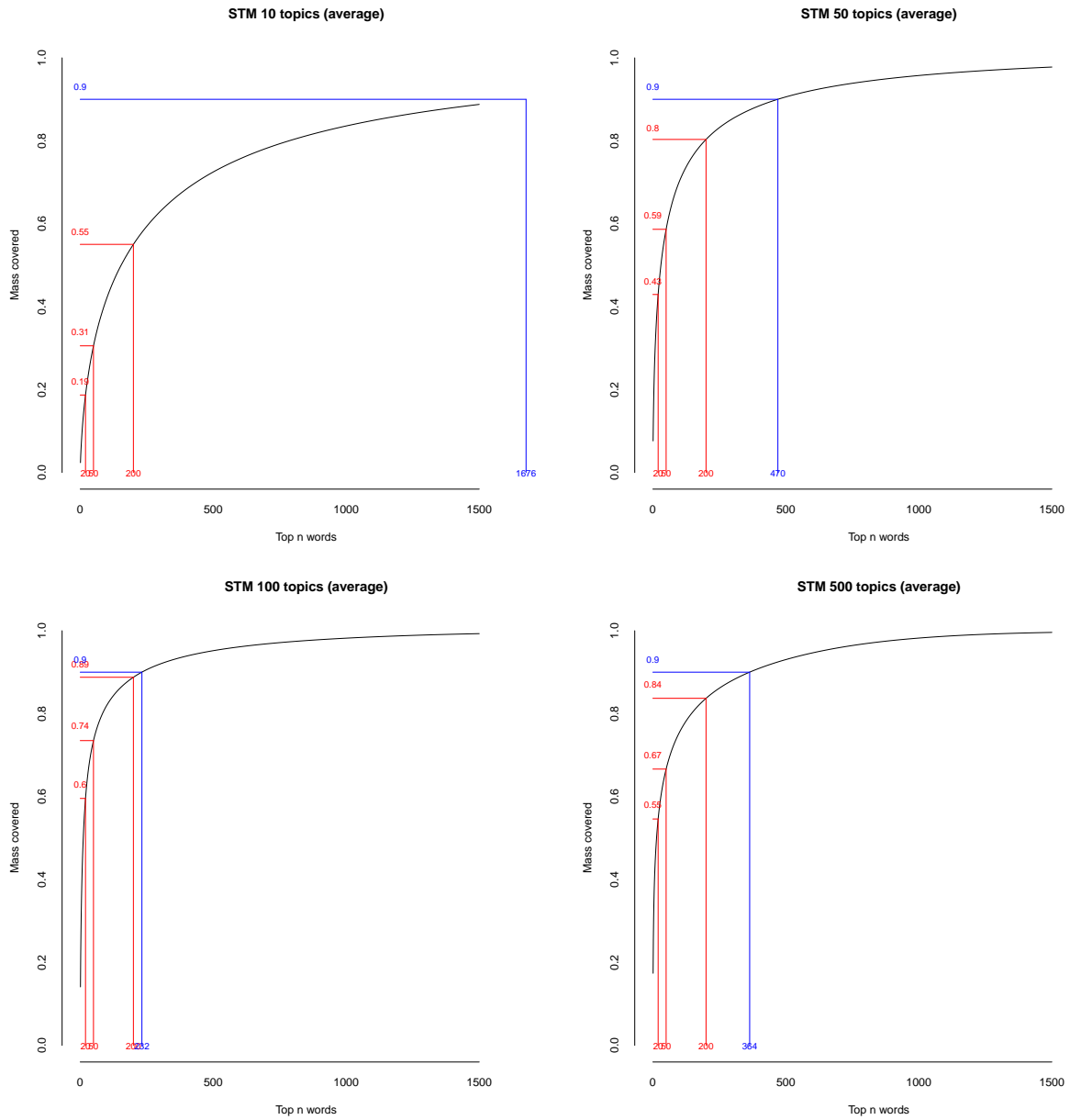


Figure SI28: Word Cloud: Six topics from **model 1** (the 10-topic model, 1 iteration)

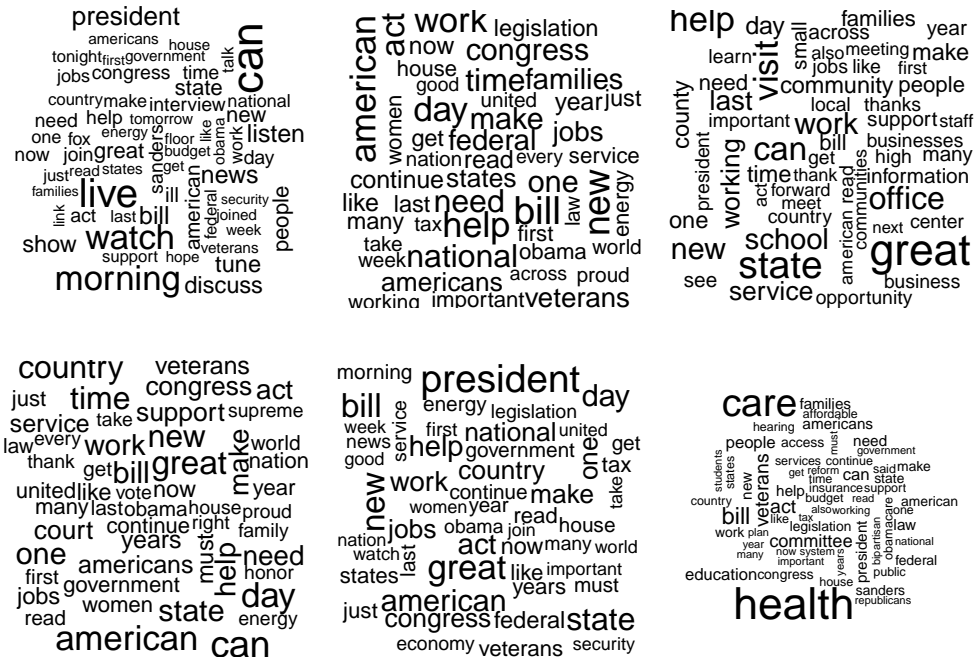


Figure SI29: Word Cloud: Six topics from **model 2** (the 10-topic model)

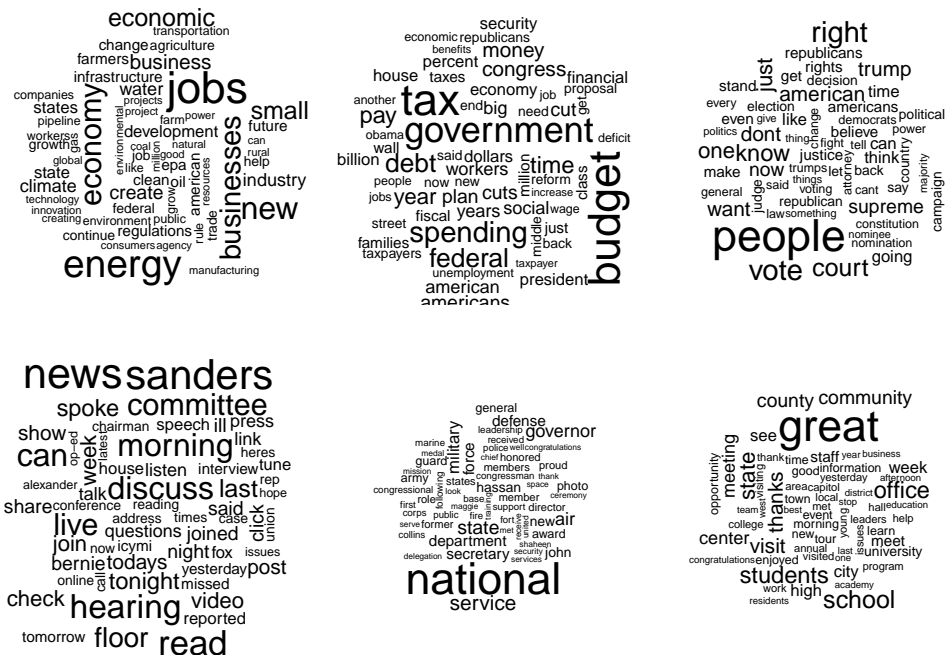


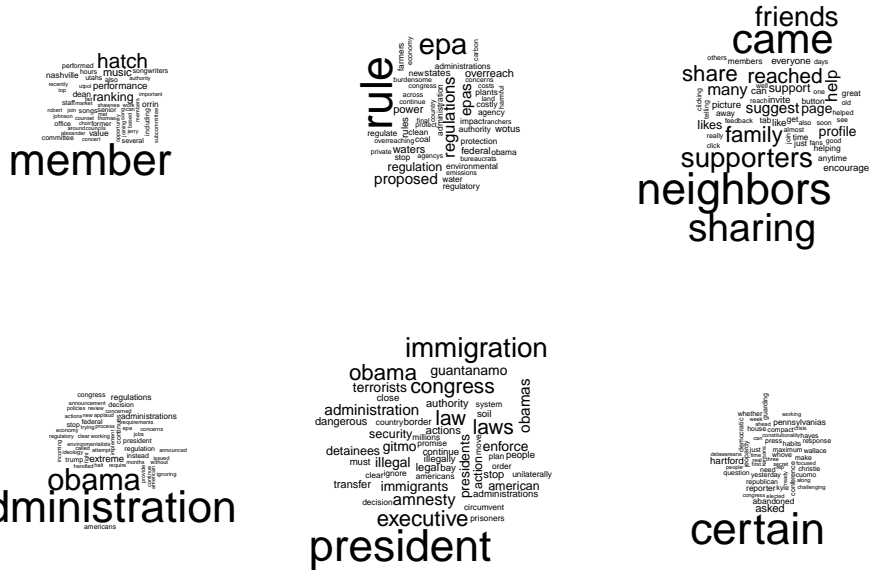
Figure SI30: Word Cloud: Six topics from model 3 (the 50-topic model)



Figure SI31: Word Cloud: Six topics from model 4 (the 100-topic model)



Figure SI32: Word Cloud: Six topics from **model 5** (the 500-topic model)



7.3 Representative Documents for Labeling

For the label validation tasks, we asked the two careful coders to read carefully the top 50 representative documents. Table SI8 shows the top 5 documents for each domestic topic with which they come up with the labels.

Table SI8: Representative Documents for Domestic Topics

Label (Careful Coder)	Document
Equal Pay for Women	(1) In Indiana, women working full-time make an average of 75 cents for every dollar earned by their male colleagues. Women in Indiana and around the country deserve to earn equal pay for equal work. When women are paid less, it not only impacts their ability to save for retirement but also means families have less to spend on groceries, rent, and other basic expenses. #EqualPayDay (2) Sweat, sweat, sweat! Work and sweat, cry and sweat, pray and sweat! Zora Neale Hurston #Hardwork

(3) Women earn less than men regardless of any other factor. I support #EqualPay for equal work & that's why I voted for the Paycheck Fairness Act that builds on the Lilly Ledbetter Fair Pay Act by closing loopholes that enable pay discrimination against women.

Share this graphic if you agree that women should earn equal pay for equal work!

(4) Today is Equal Pay Day. North Dakota women earn 73 cents for every \$1 a man earns. I support the Paycheck Fairness Act because there should be equal pay for equal work. Like/share if you agree.

(5) In Michigan, women earn 74 cents for every dollar men earn. The Gender Pay Gap doesn't just hurt women - it hurts entire families, affecting retirement, health care coverage, pensions and much more. Today is Equal Pay Day, a date that symbolizes how far into the year women must work to earn what men earned in the previous year. It is time to end the Gender Pay Gap and pay women fair wages for their work.

Healthcare/
Reproductive
Rights

(1) Planned Parenthood provides critical access to health care for thousands of Montanans. #StandWithPP

(2) Courtney from Wisconsin stands with Planned Parenthood because without them, she wouldn't have access to quality and affordable women's health care.

TrumpCare rolls back the clock on women's health by cutting funding for maternity care and defunding Planned Parenthood.

(3) I #StandwithPP for the millions who depend on their critical health care services. #PinkOut

(4) California health centers received funds to expand primary care services by hiring qualified health care providers.

(5) Planned Parenthood is providing women across Michigan with critical preventative care, including cancer screenings and wellness exams. Today, I met with Planned Parenthood Advocates of Michigan to let them know I #StandWithPP and the millions of women who rely on them for quality health care. #PinkOut

Agriculture

(1) Antibiotic overuse unchecked: "About four-fifths of all antibiotics sold in the U.S. go to livestock and poultry."

(2) Montana ranchers produce the best beef in the world and we can't have beef imports from Brazil and Argentina jeopardizing the livelihood of Montana producers.

(3) Today we recognize and thank all who contribute to Iowa's agriculture industry, among them Iowa's farmers, producers, ranchers: [Link Omitted] #AgDay

(4) Senator Leahy: #VT farmers, sugar makers and rural farms benefited from REAP in 2008 Farm Bill. #FarmBill 2012 continues this good work.

(5) Senator Shaheen discussed improvements made to the dairy program in this years farm bill during a tour of McNamara Dairy farm in Plainfield. Employees of McNamara Dairy milk 175 Holstein cows, bottle their own milk and sell it to customers up and down the Connecticut River Valley. (Plainfield, NH. July 20, 2012)

Student Loan/ Debt

(1) #ObamaEconomy: Inflation-adjusted median household income has fallen from \$54K in '08, to only \$52K in 2013 while inflation-adjusted per capita income has fallen from \$29,173 in 2008 to \$28,829

(2) Recent college graduates are struggling in this slow economic recovery. We must keep student loan interest rates at the current rate and prevent them from doubling, but we must pay for it.

(3) In our latest installment of Correspondence from the Commute, Chris replies to Bailey from Hockessin about Joseph Kony and the LRA.

(4) Americans have over \$1.3 trillion of student debt. My In #TheRedAct will allow struggling borrowers to refinance their student loans and take advantage of lower interest rates the same way people refinance a mortgage, a car loan or business debt.

(5) FACT: Total student loan debt now surpasses both credit card & auto loan debt. #HigherEdNotDebt

Drug Abuse

(1) DEA Implements Cornyn-Klobuchar Law to Help Curb Prescription Drug Abuse:

(2) Today, I teamed up with Senator Sheldon Whitehouse in hosting a roundtable with the Alliance to Prevent the Abuse of Medicines on how we can work together to combat prescription drug abuse and help struggling families and communities in Ohio. To prevent drug abuse and better help the tens of thousands of Ohioans struggling with addiction, we need a comprehensive strategy that starts from the bottom up. Our bipartisan Comprehensive Addiction and Recovery Act builds on proven methods to enable law enforcement to respond to the heroin epidemic and supports long-term recovery by connecting prevention and education efforts with treatment programs. Like my video to join me in the fight against drug abuse. [Link Omitted]

(3) Today I re-introduced the Budgeting for Opioid Addiction Treatment (LifeBOAT) Act to establish reliable funding to expand access to substance abuse treatment in West Virginia.

(4) Help prevent the misuse of addictive prescription drugs on Prescription Drug Take-Back Day at a location near you [Link Omitted]

(5) KSDK NewsChannel 5 previews Claire's roundtable discussion today at NCADA to tackle the opioid and heroin epidemic with advocates, providers, and law enforcement.

Higher Education/ Job Training	<p>(1) Very productive college affordability roundtable with students from University of Wisconsin-Parkside, Gateway Technical College, Carthage College & University of Wisconsin - Whitewater. Higher education should be a path to shared prosperity, not a path into suffocating debt.</p> <p>(2) Honored to be at Northwestern Michigan College yesterday for the expansion of their Community College Skilled Trades Equipment Program. With this new program, NMC will be able to offer wider array of educational opportunities - from marine and aviation tech to welding and nursing.</p> <p>(3) Happy to join Community College of Rhode Island (CCRI) today to announce a \$2.5 million federal grant to create new pathways and opportunities in advanced manufacturing. CCRI's Accelerated Pathways in Advanced Manufacturing program emphasizes a learn and earn model that provides opportunities for adult learners to acquire new knowledge and skills that are linked with jobs in high-growth industries.</p> <p>(4) As part of her efforts to help New Hampshire workers develop the skills and innovative thinking needed for jobs in the 21st century economy, Governor Maggie Hassan announced today that five New Hampshire companies have been awarded job training grants to help them train 171 workers in new skills. The job training grants total \$144,973.50, and the companies contributed matching funds to bring the total amount for training workers to \$289,947.</p> <p>(5) Continuing her efforts to help New Hampshire workers develop the skills, knowledge and innovative thinking necessary for jobs in the 21st century economy, Governor Hassan announced today that five companies will receive job training grants to help them train 125 workers in new skills. The job training grants total \$72,536.48. The five companies contributed matching funds, bringing the total funds for training workers to \$145,072.96.</p>
Wall Street/ Financial Sector	<p>(1) SenJohnMcCain: Wall Street bonuses: Goldman Sachs \$16.7 billion; JP Morgan \$8.7 billion; main street \$0 your tax doll...</p> <p>(2) This recession was caused by the greed, the recklessness and the illegal behavior on Wall Street. [Link Omitted]</p> <p>(3) Will question CEO Jamie Dimon about J.P. Morgan's \$2 billion plus losses at a Senate Banking hearing around 11am. Tune in here: [Link Omitted]</p> <p>(4) I'm working with Consumer Financial Protection Bureau head Rich Cordray to protect consumers from predatory for-profit schools & big bank tricks.</p> <p>(5) No, AIG. Criticism of Wall Street bonuses is not equivalent to lynchings. Offensive.</p>
Government Shutdown/ Congressional Budget	<p>(1) Republicans in Congress are passing bills to reopen and fund essential functions of our government. It's the Democrats in Congress who refuse to negotiate. #HarryReidsShutdown</p>

(2) How does federal government claim preemption in Arizona when federal government fails to act?

(3) This week, Puerto Rico argued before the Supreme Court that it should be able to restructure its debt. If Puerto Rico were a country, it could go to the IMF. If it were a city, it could declare bankruptcy. But because Puerto Rico is a U.S. territory, it can do neither. Without help, Puerto Rico is at the mercy of Wall Street vulture funds, creditors that specialize in preying on borrowers who are in trouble. The Republican leadership needs to step in immediately and help millions of American families who are caught in an economic catastrophe.

(4) Skipping the budgeting, authorizing and appropriating committee process in lieu of last-minute, omnibus, must-pass, stop-gap bills diminishes the role of Congress. It diminishes the role of individual members of Congress and reduces the input of the people from that members state or district. By skipping steps, Congress allows the Administration to regulate and spend with little fear of being checked by the legislative branch. Our countrys founders gave Congress the power of the purse. The Senate and House should right the ship no matter what party controls the White House. Using the committee system and considering legislation as it should be considered would go a long way. It would put a stop to a lot of partisanship, some of which is simply frustration from leaders preventing members from legislating.

(5) An essential premise of good government is that Congress should authorize programs and activities before it funds them. If we relinquish our responsibility to regularly review and reform these programs, all of our government funding will essentially operate on auto-pay.

Obamacare/
Tax Policy

(1) #99countymeetings 50ppl at Mapleton Q&A EPAregs VoterID MedicareFraud Obamacare/SupremeCt Farmbill F&F ElectionFinanceRecorm SocSec

(2) Courtesy of ObamaCare: Broken promises, plagued by failure & cost skyrocket for millions.

(3) #ObamaCare creates fewer options at “much higher premiums, higher out-of-pocket costs, higher taxes on the costs that they[everyday Americans] do incur, and fewer jobs and fewer hours for those who are employed.”

(4) Millions who lost their insurance and reenrolled through OCare are facing higher premiums: [Link Omitted]

(5) Premiums are soaring, patients choices are dwindling. Obamacare must be repealed and replaced. #RepealObamacare

Deficits/
Debt/ Budget

(1) The President’s budget proposes deficits of \$5.3 trillion and \$8.1 trillion in new debt over the next decade. Where’s the Balanced Budget?

(2) Why we must cut spending and why simply raising taxes won’t solve our fiscal problems

(3) Today the Congressional Budget Office released a report that shows positive impact of the budget caps. If our nation is serious about balancing our budget and reducing Americas debt, real, substantive budget reforms and savings will have to be on the table during any spending negotiations.

(4) Voted NO on raising debt ceiling without addressing spending problem. Need to both prevent a debt limit crisis today & debt crisis tomorrow.

(5) The combination of new spending with the President's proposed budget freeze will only maintain an unsustainable status quo. We need serious cuts, not a continuation of the trillion dollar deficits that are jeopardizing our nations fiscal stability.

Table SI9 shows the top 5 documents for each domestic topic with which they come up with the labels.

Table SI9: Representative Documents for International Topics

Label (Careful Coder)	Document
International Trade	<p>(1) Sherrod reintroduces China trade plan to discourage currency manipulation that harms American exports. As our trade deficit continues to widen, our need to level the playing field for American manufacturers and workers becomes more urgent.</p> <p>(2) RT @GrassleyOffice: Heres #TheScoop by @ChuckGrassley [Link Omitted] #4jobs @RedCross @Mail4Heroes #fastandfurious #NationalGuard ...</p> <p>(3) Chinas economy is simply too large for it to be artificially propped up by a blatantly manipulated Yuan. Chinas actions are deliberate and are designed to give China a competitive advantage in the marketplace. I believe free and fair trade is beneficial, but I also know that Chinese currency manipulation and intellectual property theft is doing harm to our economy.</p> <p>(4) I am pleased that the U.S. Trade Representative took action against Chinas unfair trade practices. Its policies give Chinese companies an unfair advantage and hurt American workers and products.</p> <p>(5) Today, I urged the International Trade Commission to protect Ohio Steelworkers. Ohio has the best steelworkers in the country and they deserve to compete on a level playing field with our competitors. That is why I am working to promote exports of Ohio steel and protect our steelworkers from unfair competition: [Link Omitted]</p>
Praising Ac- tive Military/ Military Units	<p>(1) 150 soldiers from the 1st Stryker Brigade Combat Team, 25th Infantry Division arrived home to Fort Wainwright yesterday from a deployment in Afghanistan. Welcome home! Check out these great homecoming pictures:</p>

(2) The South Dakota Army National Guard 196th Maneuver Enhancement Brigade returned after a 10-month deployment to Kuwait. I was honored to welcome them home today.

(3) Welcome support from Hollywood to #SaveSaeed Abedini and Kenneth Bae.

(4) Welcome home from Afghanistan SPC Jessica McKim, The U.S. Army!

(5) Yesterday, the 82nd Airborne Division welcomed its 49th commander, Maj. Gen. Erik Kurilla. Join me in thanking Maj. Gen. Kurilla and the 48th commander, Maj. Gen. Richard Clarke for their service to the 82nd Airborne Division.

Terrorism

(1) Whether its in Jerusalem, or Paris, or New York, the civilized people of the world are under siege by violent Islamic extremists. We must redouble our efforts to win this war on terrorism.

In 2016, we must remain vigilant against Islamic extremism. We need to do everything in our power to protect Americans from terrorism.

(2) Washington must wake up to this fact: the crowd-sourcing of domestic terror is a reality. While slow-moving, federal bureaucracies look for card-carrying terrorists, the Islamic State and al Qaeda are crowd-sourcing their jihad. We must adapt and develop a long-term strategy to name and defeat our enemy. That enemy is not the empty label “extremism” but the ideology of militant Islam.

(3) This weekend as Americans celebrate the birth of our nation and our freedoms, the horrific events yesterday are a sobering reminder of the radical Islamic terrorists who seek to destroy us. Twenty more innocent civilians are dead in the most recent attack by radical Islamic terrorism this time the target was an upscale neighborhood in Dhaka, #Bangladesh. Terrorists throwing grenades and shouting Allahu Akbar attacked a popular restaurant frequented by diplomats and students, reportedly slitting the throats of some of their victims and singling out those who could not recite verses from the Koran.

Over the past few years, the authorities in Bangladesh have insisted that terrorist incidents in their country were somehow isolated from the global scourge of radical Islamic terrorism, the product of homegrown militants responding to domestic grievances. This attack, however, appears designed by ISIS to prove otherwise, and that the Islamic State is functioning not only in Syria and Iraq but also in Bangladesh, just as it is in Turkey, France, Belgium and the United States.

This isn't just a wake-up call anymore it is a screaming siren of an alarm that demands our urgent attention. Just as the Bangladesh attack was a demonstration of the Islamic States determination to take their war on civilization to another Muslim-majority nation, the recent attacks in the United States from San Bernardino to Orlando demonstrate the United States is, likewise, not immune. We are fortunate to have friends and allies in Europe, the Middle East, and Asia who see this enemy for what it is, and we should do everything we can to partner with them not to struggle against generic violent extremism but to actually take the fight to the radical Islamic terrorists.

(4) The attack at the Istanbul Ataturk Airport in Turkey is yet another tragic reminder of the threats we face from terrorists who indiscriminately employ senseless violence with no regard for innocent lives or their own. It's a reminder that we must remain vigilant not only here at home, but also around the globe in our efforts to prevent terrorist threats at our transportation hubs and other vulnerable locations.

(5) While this is a milestone that we have all awaited, we must remember that al Qaeda and its affiliates are not dependent on one man and we must remain vigilant in our efforts to disrupt and destroy terrorist networks that threaten our Nation and allies.

Military Sexual Assault

(1) Momentum continues to build in support of the Military Justice Improvement Act, which would create an unbiased military justice system by transferring the decision-making authority over which cases go to trial from the chain of command to independent trained military prosecutors where it belongs. I hope you'll join me and my colleagues in supporting this commonsense measure to reform the way the military handles sexual assault cases so that victims can get the fair shot at justice they deserve.

(2) Military sexual assault victims were silenced in Congress. Lend your voice in support of the Military Justice Improvement Act to give them a fair shot at justice. #MJIA

(3) Yesterday, I stood alongside sexual assault survivors, advocates, and a bipartisan coalition of U.S. Senators to renew our push for real reform of the military justice system. Our bipartisan Military Justice Improvement Act would remove prosecutorial authority in sexual assault cases from the chain of command and place it in the hands of independent trained military prosecutors where it belongs. Only then will we truly create an unbiased system of military justice where assailants are held accountable and survivors, like Samantha Jackson, get the justice they deserve.

(4) Of the estimated 26,000 incidents of unwanted sexual contact and sexual assaults that occurred in 2012, only 3,374 were reported. Time and again, victims tell us the reason there is such a drop off in reporting is because the decision-making in cases of sexual assault lies within the chain of command. Until this authority is moved outside the chain of command, into the hands of an independent military prosecutor, victims will not have the confidence they will receive the justice they deserve. This is why I'm fighting to pass the Military Justice Improvement Act, to make sure the voices of victims—like Navy veteran and MST survivor Brian Lewis—are heard and that we create an objective and accountable military justice system. Please share this graphic to let Brian's voice, and that of so many others, be heard. #MJIA

(5) Victims of sexual assault in the military tell us time and time again that the reason they don't report sexual assaults is because the decisionmaking lies within their own chain of command. If we want to truly reform the military justice system and give victims of sexual assault a real chance at justice, we must remove this decisionmaking power from the chain of command and place it in the hands of independent trained military prosecutors. Right now, victims' voices are not being heard in Congress, such as U.S. Marine Vet & #MST survivor Stacey Thompson. Please lend your voice in support of the Military Justice Improvement Act and make sure the voices of victims like Stacey are heard: [Link Omitted] #MJIA

Nuclear De-
terrence/
International
Security

(1) Today, Iran test-fired ballistic missiles in violation of international sanctions. This launch demonstrates the regimes continued disregard for international sanctions and the grave consequences of this Administration blindly trusting the Iranian regime to uphold its commitments. Instead of making concessions to the Iranian regime and freeing up more than \$100 billion dollars in sanctions relief, the Obama Administration should change course and increase pressure on Iran by imposing additional unilateral sanctions to stop their belligerent behavior once and for all.

(2) The more we learn about the Iran deal, the more concerned I am that it not only fails to prevent Iran from obtaining a nuclear weapon, it emboldens them through tens of billions of dollars in sanctions relief, a phased out lifting of a UN arms embargo and the ability to test more advanced centrifuges.

(3) The long list of concessions in the Presidents Iran deal includes the end of a long-standing arms embargo. This misguided concession, coupled with the immediate and imprudent sanctions relief, makes this not only a weak deal, but a dangerous one, too.

(4) Even President Obama is admitting that Iran would be able to develop a nuclear weapon in “a matter of months” after the Iran deal expires. I strongly oppose the deal, which paves the path for Iran to become a nuclear power and threatens the security of Israel and our partners in the Middle East.

(5) Freeing American prisoners of conscience in Iranian prisons should be a pre-condition to any negotiations with Iran. #FreeAmir

Air Force

(1) Today, Brig. Udo K. “Karl” McGregor, commander of the 452nd Air Mobility Wing, March ARB, Calif., passed the 730th Air Mobility Training Squadron guidon to Lt. Col. Jonathan M. Philebaum, 730th AMS commander, Altus AFB, during the squadron reactivation and assumption of command at hangar 517. (Photo courtesy of U.S. Air Force)

(2) It was an honor to spend Friday at Pease Air National Guard Base, onboard the KC-135A Stratotanker with the New Hampshire National Guard. During the refueling mission, I watched as the crew honed in on a EC-130J Commando Solo aircraft not far from Pittsburgh and, with laser like precision, guided a refueling hose into the aircraft.

(3) Advocated for Warriors of the North-Grand Forks Air Force Base, Minot Air Force Base, and North Dakota National Guard during meeting with U.S.Airforce.

(4) U.S. Sen. Cory Booker and the Adjutant General of the New Jersey National Guard, Brig. Gen. Michael L. Cunniff, discuss the flight parameters of an F-16 fighter jet during a visit to the 177th Fighter Wing in Egg Harbor Township, N.J. on Aug. 5, 2016. It was the first visit to the New Jersey Air National Guard base by Sen. Booker, who received a briefing by the Wing Commander, U.S. Air Force Col. John R. DiDonna, flew an F-16 simulator and got a chance to get an up close look at an F-16C Fighting Falcon fighter jet on the flight line at the Atlantic City Air National Guard base. (U.S. Air National Guard photo by Master Sgt. Andrew J. Moseley/Released)

(5) Enjoyed meeting with Col. Gentry Boswell, Commander of the 28th Bomb Wing at Ellsworth Air Force Base and Command Chief, CMSgt Sonia Lee today to discuss the status of the B-1 fleet at Ellsworth.

Honoring
Specific
Veterans

(1) What an honor to present the Distinguished Flying Cross to WWII Veteran and Hudson-area resident Alfred LeFeber. A long overdue recognition for a true patriot.

(2) Yesterday, I was honored to present Roger Piasecki of Kearney with a WWII Medal & Lapel Pin for his father Walter’s military service. #Greatest-Generation

(3) Today I participated in a ceremony honoring Mildred Pretzer of Charlottesville, a World War II veteran who served as part of the Womens Army Corps (WAC). I proudly presented Ms. Pretzer with the Army Good Conduct Medal, the WAC Service Medal, the American Campaign Medal, the World War II Victory Medal and the Honorable Service Lapel Pin.

(4) Today, I had the privilege of escorting James Gray, a WWII Veteran from Malvern as he visited the WWII Memorial for the first time. Mr. Gray is a former POW and Purple Heart Recipient who fought in the Battle of the Bulge where he was captured and spent six months as a German POW. Thank you for your service to our great country, Mr. Gray!

(5) It was an honor to present lost medals to World War II Veteran and hero Julious Elmore in Magnolia today.

Honoring
Veterans/
Heroes

(1) Honor the heroes who ran into burning buildings. Honor the brave who sacrificed their lives for strangers. And never forget those lost on 9/11/2001 and 9/11/2012.

(2) Today we remember and honor the brave men and women that sacrificed their lives for our freedom. We are forever grateful and you will never be forgotten.

(3) Thank you to the men and women who have worn, and continue to wear, the uniform of our nation's military. As America pauses today to remember the service and sacrifice that our veterans and active duty members have endured, we are forever in debt to you and your families for your selfless service.

(4) Today, let us pause to remember the heroes we have lost in the service of our great nation. Thank you to all who serve to protect our freedom.

(5) This #MemorialDay, we honor the brave men and women who paid the ultimate sacrifice to preserve our nation's freedom

Military
Operations/
Armed Con-
flicts

(1) Sen. Corker visited the Kilis refugee camp on the Turkish border with Syria today for the second time since the conflict began. Members of the camps refugee leadership expressed dismay and disappointment at the lack of U.S. support for the Syrian opposition. Specifically, refugee leaders noted the importance of helping arm and equip the vetted, moderate opposition under the leadership of Gen. Salim Idriss.

(2) Congressional authorization for the use of military force against ISIS, Al-Qaeda, and the Taliban will make clear to our warfighters, our allies, and our adversaries that we are united #AUMF

(3) U.S. Senator Tim Kaine leads a Congressional Delegation (CODEL) to the Middle East region focused on the U.S. mission against ISIL and the ongoing humanitarian crisis in Iraq and Syria.

(4) The drawdown of US troops in Iraq and use of force in Libya without a plan have enabled ISIS to amass power and territory with little pushback.

(5) ISIS has possession of and is now using illegal chemical weapons against the Kurds in Iraq. Earlier this year, I visited Iraq and met with the leaders of Iraqi Kurdistan, and I firmly believe that the U.S. must increase its support for the Iraqi Kurdish Peshmerga forces who are valiantly fighting against ISIS.

Veterans
Affairs/
Veterans
Healthcare

(1) I have been and will continue to be committed to reforming the VA to ensure our veterans receive the care they deserve.

For our veterans in need of help with the VA, my office is here to help!

(2) Senator Grassley has made it clear that the comments from the Secretary of Veterans Affairs comparing wait times for VA treatment with wait times for rides at Disneyland were unacceptable and that the Secretary should make amends.

Senator Grassley has worked to improve veterans experience at the VA, including pushing for and receiving an acknowledgement from the VA that veterans have experienced problems with the Veterans Choice program and a pledge to improve services to Iowa veterans and veterans across the country.

(3) “Our veterans deserve the best medical care available, but far too often, veterans suffering from PTSD, Traumatic Brain Injury, and other psychological impacts slip through the cracks.”

(4) The Department of Veterans Affairs needs to fix problems in helping veterans get appointments outside the VA when needed. Senator Grassley is working with the VA on veterans’ frustrations.

(5) Today, I toured the Portland Vet Center with Steven Reeves from the U.S. Department of Veterans Affairs. Our veterans deserve the highest quality care possible, and our 5 Vet Centers in Maine need to be adequately staffed in order to provide vital services to those who served our country.

References

- Barberá, P., A. Casas, J. Nagler, P. J. Egan, R. Bonneau, J. T. Jost, and J. A. Tucker (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review* 113(4), 883–901.
- Blaydes, L., J. Grimmer, and A. McQueen (2018). Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds. *Journal of Politics* 80(4), 1150–1167.
- Chang, J., S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems*, pp. 288–296.
- Dietrich, B. J., M. Hayes, and D. Z. O’Brien (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review* 113(4), 941–962.
- Mimno, D., H. M. Wallach, E. Talley, M. Leenders, and A. McCallum (2011). Optimizing Semantic Coherence in Topic Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 262–272. Association for Computational Linguistics.
- Nielsen, R. A. (2020). Women’s Authority in Patriarchal Social Movements: The Case of Female Salafi Preachers. *American Journal of Political Science* 64(1), 52–66.
- Pan, J. and K. Chen (2018). Concealing corruption: How chinese officials distort upward reporting of online grievances. *American Political Science Review* 112(3), 602–620.
- Roberts, M. E., B. M. Stewart, and E. M. Airoldi (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association* 111(515), 988–1003.
- Roberts, M. E., B. M. Stewart, and R. A. Nielsen (2020). Adjusting for confounding with text matching. *American Journal of Political Science* 64(4), 887–903.
- Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand (2014). Structural Topic Models for Open-ended Survey Responses. *American Journal of Political Science* 58(4), 1064–1082.
- Rozenas, A. and D. Stukal (2019, June). How Autocrats Manipulate Economic News: Evidence from Russia’s State-Controlled Television. *The Journal of Politics* 81(3), 982–996.
- Wallach, H. M., I. Murray, R. Salakhutdinov, and D. Mimno (2009). Evaluation Methods for Topic Models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1105–1112. ACM.