Estimating the Prevalence of Religious Content in Intelligent Design Social Media

George D. Montañez
Machine Learning Department
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
gmontane@cs.cmu.edu

***Abstract*—Can machine learning prove useful in deciding sociological questions that are difficult for humans to judge impartially? We propose that it can, and even simple methods can be useful for evaluating evidence with reduced influence from human bias. Our case study is intelligent design (ID) social media, particularly the detection of religious content therein. Being a polarizing topic, critics of intelligent design claim that all intelligent design output consists of religious content, whereas defenders argue that ID is primarily motivated by scientific, not religious, concerns. To help determine where the truth lies, we use classifiers trained on the topically categorized 20 newsgroups dataset, applying the trained learners to automatically classify ID blog documents. As a control, we perform the same analysis on documents drawn from prominent mainstream evolutionary science blogs. Our classification results demonstrate a significant portion of religious and political content in the intelligent design dataset as judged by a non-human classifier, and a similarity in the proportion of documents assigned to religious and political categories in the evolutionary science blog dataset, likely indicating a dependence of discussion topics within the two communities.**

I. Introduction

Sometimes "I know it when I see it" just isn't good enough. Human bias is an ever present threat when making social judgements and performing research [1], [2], [3]. When asking questions about the presence of offensive content, the disposition (and experiences) of a human observer can affect whether such content is present. This is to say nothing of malicious users, who incorrectly report content to further personal or group goals [4]. Thus, having humans subjectively estimate whether proscribed content is present in a document often results in conflicting answers. Researchers have long realized the potential of automating content judgements [5], [6], [7], [8], [9], [10], [11]. Automated detection of proscribed content in digital sources has been explored for such domains as spam email [5], [6], [7], malware detection [12], [13], ad hominem ("flame") detection [8], [9], [10], and fake reviews [14], [15], [11], [16]. Inspired by this work, we use automated methods to estimate the proportion of religious content in presumably non-religious sources, with the goal of reducing the influence of subjective bias in forming such estimates.

Our study focuses on the detection of religious and political content in documents drawn from prominent *intelligent design* (ID) blogs. Most scientists dismiss intelligent design as a form of religious creationism, backed by an American movement seeking to establish a conservative political agenda [17]. Proponents disagree, claiming scientific motives separate from religious or political concerns [18], [19], [20]. Disagreement, therefore, exists concerning the expected amount of religious and political content in intelligent design material. Through the use of automated text classification and probabilistic topic modeling, we seek to empirically estimate the actual proportion of such content in intelligent design documents, thus reducing the bias present in human judgements.

Our goal in this study is not to introduce novel or complex machine learning techniques, but to show that even simple methods, such as standard naïve Bayes classifiers, can be useful in this regard. We use a naïve Bayes classifier trained on the well-known 20 newsgroups dataset [21], which contains class categories corresponding to scientific, political and religious newsgroups. Using the trained classifier, the documents in our datasets are classified into their respective categories, with the proportion of ID documents assigned to religious and political categories serving as our approximate (and hopefully less biased) measure of religious and political content in ID blogs. For the measure to have validity, the chosen classifier must have high classification accuracy on the 20 newsgroups dataset, as well as have high classification accuracy on non-20 newsgroups documents drawn from blogs for which class categories are known. We therefore propose to test the following four hypotheses:

1) Under cross-validation, the accuracy of the classification method should be high for documents within the 20 newsgroups dataset.
2) For (non-20 newsgroups) documents whose natural group categories are clear, the classification method should be highly accurate in assigning new documents to their expected categories.
3) Documents in the ID document set are more likely to be assigned to religious and political newsgroups than to scientific newsgroups.
4) Statistically significant differences should exist between documents drawn from the ID document set and those drawn from mainstream evolutionary science blogs in the percentages of documents assigned to religious and political categories; ID documents should be assigned to religious and political newsgroups more frequently than documents drawn from mainstream evolutionary science blogs.

The first two of these hypotheses serve as a test of the accuracy of our method on ground truth data, and the final hypotheses serves as a control, to put into perspective any insights gained from our analysis of the ID dataset. The third hypothesis is the primary question under investigation and the motivation for studies of this kind.

We begin with a brief review of naïve Bayes classification and present an augmented feature representation using local co-occurrence of word pairs, before evaluating each of the above four hypotheses in turn. Though naïve Bayes classification is quite simple, we show the high accuracy of naïve Bayes classification on the 20 newsgroups dataset under our featurization, as well as high classification accuracy on non-20 newsgroups blog documents for which class categories are known. In assessing the intelligent design dataset, our method suggests a significant proportion of religious and political content in ID blogs, while unexpectedly revealing a similar distribution of religious and political content in evolutionary science blogs, likely indicating an interdependence between the two datasets. We conclude with a brief discussion of the results and some caveats concerning their interpretability.

II. Naïve Bayes Classification

Naïve Bayes classification is a machine learning method based on Bayes' Theorem under the assumption of conditional independence of features given a class [22]. For text classification, we train on a corpus of documents to estimate the posterior probability of the class given a document, which is proportional to the probability of the document given a class multiplied by the probability of the class, namely

$$p(l \mid d) \propto p(d|l)p(l) = p(l)\prod_{w} p(w|l)$$

where l is the class, d is the document and each w is a word in the document. The conditional independence assumption allows the conditional probability of the document given the class to be represented as the product of the conditional likelihoods for the individual words, which greatly reduces the amount of data needed for parameter estimation. Given a trained classifier and a document to be classified, the classifier outputs the class label l_{NB} that maximizes

$$l_{\text{NB}} = \operatorname*{argmax}_{l_j \in \mathcal{L}} p(l_j)\prod_{w} p(w|l_j)$$

where l_{NB} is the predicted class, $\mathcal{L}$ is the set of possible classes and each w is a word in the document.

Naïve Bayes classification makes the relatively strong assumption that given a class label, features in a document (such as words) are independent of one another. While this assumption is often wrong, it has been shown that naïve Bayes classification can work well even in problem settings where the conditional independence assumption is known to be violated [22]. While violation of the conditional independence assumption leads to errors in posterior probability calculations, the zero-one loss classification error remains relatively unaffected, since relative orderings of class probabilities remain unchanged. Thus, naïve Bayes classification has enjoyed wide success in a variety of domains, including text-classification [21].

A. *Naïve Bayes Classification with Local Co-occurrence (LCO) Features*

For naïve Bayes classification of text, an often used featurization is to represent each document as a simple "bag-of-words", which consists of a list of terms and their frequency of occurrence within the document. Here we consider an extension to the unigram bag-of-words representation that incorporates *local co-occurrence* (LCO) features, which are counts of word pairs occurring together in sentences, though not necessarily adjacent to one another in a given sentence, as in standard n-grams. By allowing for variable skip-lengths, these can be seen as a form of skip-gram [23] with a variable skip-length determined by period position. We restrict sentences to those with fifty words or less, ignoring those larger than this when learning LCO features. By bounding the maximum sentence size, the learning process remains linear in the number of words in the corpus, though having a greater constant factor.

As is the case with bigrams, the additional LCO pairs can add up to $|V|^2$ new features, where $|V|$ is the size of the classifier vocabulary. To reduce the number of LCO features and their effect on the final posterior probabilities, we introduce two additional parameters, c, which establishes a minimum occurrence frequency for word pairs before they are incorporated into the model, and α, a weighting parameter that controls the contribution of the LCO features to the overall log-likelihood. This leads to an expanded probabilistic model for our naïve Bayes classifier with LCO features, represented as

$$p(l \mid d) \propto p(l) \prod_{w} p(w \mid l) \prod_{(w_1,w_2)} p((w_1,w_2) \mid l)^{\alpha}$$

where l is the class, d is a document, w is a word in the document, α is the (log) weight for LCO features and (w_1, w_2) are LCO word pairs occurring at least c times in the document.

For a wide variety of parameter settings tested, the LCO naïve Bayes performed as well as or better than the non-LCO model, with statistically significant improvements in classification accuracy on the 20 newsgroups dataset.

III. 20 Newsgroups Experiments

Four datasets are used for our analysis, the first being the widely used 20 newsgroups dataset [21], and the additional three being collected for the purpose of this study. All documents were preprocessed to remove non-alphanumeric symbols, made lowercase and documents with fewer than forty words were omitted.

The 20 newsgroups dataset consists of roughly eighteen-thousand distinct documents collected from twenty different usenet newsgroups during the late nineteen-nineties [21]. For this study, we used J. Rennie's reduced 20 newsgroups dataset [24], which consists of 18,828 individual documents. After removing email addresses, non-alphanumeric symbols, "From:" document headers and omitting documents with fewer than forty words, we retained a set of 17,768 documents. Each of these documents belong to exactly one newsgroup from the following: comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, comp.sys.mac.hardware, comp.windows.x, rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey, sci.crypt, sci.electronics, sci.med, sci.space,

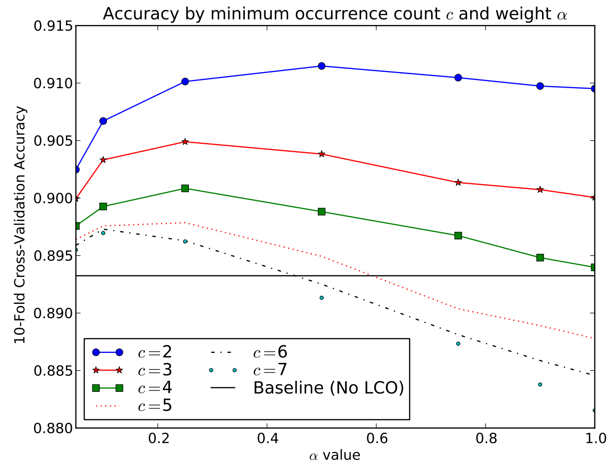

Fig. 1. 10-fold cross-validation accuracy for several parameterizations of naïve Bayes with LCO.

misc.forsale, talk.politics.misc, talk.politics.guns, talk.politics.mideast, talk.religion.misc, alt.atheism, and soc.religion.christian.

We tested the classification accuracy of naïve Bayes on the 20 newsgroups dataset using 10-fold cross-validation, both with and without the additional LCO features. In both cases, we removed common English stopwords using the standard SMART stoplist and used plus-one pseudocount smoothing for word frequency counts. We used zero-one loss to measure accuracy, counting all incorrect newsgroup assignments as equal, regardless of which newsgroup a document was incorrectly assigned to.

A. Results: 20 Newsgroups

The basic naïve Bayes classifier (without LCO features) achieves a 10-fold cross-validation accuracy of $0.893 \pm .005$ (95% confidence level) on the 20 newsgroups dataset, compared with the 5.3% base rate using the most common class. The classification accuracy improves to $0.911 \pm .004$ (95% confidence level) over the same dataset when using naïve Bayes with LCO features ($c = 2, \alpha = .5$). The improvement is statistically significant at the .05 level, using a Kolmogorov-Smirnov test and applying a Benjamini-Yekutieli multiple hypothesis test adjustment [25]. Figure 1 shows 10-fold cross-validation accuracies for several different parameterizations of the naïve Bayes LCO classifier.

In summary, we see that the naïve Bayes classifier using LCO features has high accuracy on the 20 newsgroups dataset, supporting our first hypothesis.

TABLE I
10-FOLD CLASSIFICATION RESULTS FOR NON-20 NEWSGROUPS DATASET

Blog	Acc.	95% CI	# of Docs
insidethemiddleeast	0.899	±0.026	527
bats	0.959	±0.009	1,806
thegospelcoalition.org	0.911	±0.015	1,404
space	0.976	±0.014	495
crypto.com	0.782	±0.109	55
Combined (all five)	0.936	±0.007	4,287

IV. NON-20 NEWSGROUPS EXPERIMENTS

We next assessed the classification accuracy of the 20 newsgroups trained naïve Bayes classifier on non-20 newsgroups documents drawn from the web. We collected a set of 4,529 documents from five blogs, which correspond to five different newsgroup categories: bats.blogs.nytimes.com (1,806 documents), newscientist.com/section/space (495 documents), crypto.com (55 documents), thegospelcoalition.org/blogs/tgc/ (1,404 documents), and insidethemiddleeast.blogs.cnn.com (527 documents). These documents had natural classification labels, namely rec.sport.baseball, sci.space, sci.crypt, soc.religion.christian, and talk.politics.mideast, respectively. Documents from the first four blogs were downloaded from the web on March 16th-17th, 2012 and the documents from insidethemiddleeast.blogs.cnn.com were downloaded on March 24th, 2012. Documents in this set were preprocessed in the same manner as the 20 newsgroups documents.

A. Results: Non-20 Newsgroups

The LCO naïve Bayes classifier achieves a cross-validation accuracy of 0.936 ± 0.007 (95% confidence level) on the non-20 newsgroups blog document set. Table I shows the results for the individual blogs in the set as well as the combined result. The classifier achieves high classification accuracy over all blogs in the set, with the lowest accuracy above 78% (cf. base rate of 5%). Thus, a 20 newsgroups trained classifier can be used to classify at least some non-20 newsgroups documents with high accuracy, supporting our second hypothesis.

V. INTELLIGENT DESIGN BLOG ANALYSIS

Having found support for our first two hypotheses, we now discuss the set of documents drawn from intelligent design blogs, as well as compare them to a similar set of documents drawn from evolutionary science blogs. The ID dataset consists of 18,739 documents collected from the ten most prominent ID blogs: Uncommon Descent (7,892 documents), Evolution News and Views (3,546 documents), Telic Thoughts (2,175 documents), Access Research Network (ARN) (3,066 documents), Biologic Institute blog (21 documents), ID in the UK (184 documents), Intelligently Sequenced (481 documents), Intelligent Reasoning (636 documents), Research on ID (87 documents), and Darwin's God (651 documents). The evolutionary science blog dataset consists of 12,032 documents collected from ten prominent evolutionary science related blogs: Panda's Thumb (4,188 documents), Pharyngula (596 documents), Why Evolution is True (3,029 documents), NCSE (National Center for Science Education) blog (1,367 documents), ERV (821 documents), The Loom (669 documents), Talk.Origins (167 documents), Evolution (74 documents), Evolutionblog (1,110 documents) and Sandwalk (3,454 documents).

Documents from these blogs were downloaded from the web on March 16th-17th, 2012 with the exception of documents from Telic Thoughts and the Sandwalk, which were downloaded on March 24th and May 5th, 2012, respectively. All documents were preprocessed in the same manner as the 20 newsgroups dataset. The scraping software downloaded all documents available, from the initial posts onwards, thus forming a dataset showing the state of discourse in 2012 (seven years after the conclusion of the Dover trial [26], [27]), and the evolution of that discourse from the founding of the blogs. Both datasets will be made publicly available, to aid other researchers in investigating the historical progression of intelligent design and evolutionary discussion in English-language social media during this time period.

A. Methods

To classify the ID documents, an LCO naïve Bayes classifier was first trained on the full set of 20 newsgroups documents, using LCO parameters of $c = 2$ and $\alpha = 0.5$. The newsgroup classes were split among five separate category types (science, religion, politics, atheism and other) as shown in Table II. For example, documents were counted in the "science" category if classified as either sci.med, sci.crypt, sci.space, or sci.electronics, and counted as religious if assigned to talk.religion.misc or soc.religion.christian. Although an inclusive definition of religion would lead us to include alt.atheism in the religion category [28], we separate it for two reasons. First, many atheists would strongly disagree with the characterization of atheism as a religion and second, even if atheism is viewed as a form of religious expression, it is not the type of religion defenders of intelligent design are likely to espouse.

TABLE II
NEWSGROUP CATEGORY MAPPINGS

Category	Newsgroups
Science	sci.med, sci.crypt, sci.space, sci.electronics
Religion	talk.religion.misc, soc.religion.christian
Atheism	alt.atheism
Politics	talk.politics.misc, talk.politics.guns, talk.politics.mideast
Other	Remaining ten newsgroups

Thus, we test for assignment to a separate atheism category, allowing one to either include or exclude it from final category counts.

B. Results

Tables III and IV show the per blog classification results by group type percentages, as well as the overall results for the entire datasets. Figures 2-4 show the same results in graphical form, with bar width representing percentage of documents assigned to each category. The percentage of ID documents assigned to scientific newsgroups is 32.7% (±0.7%), while 15.2% (±0.5%) are assigned to the religious category (excluding atheism) and 16.2% (±0.5%) are assigned to the political category. A significant portion (roughly one-third of the documents), are therefore assigned to a religious or political category, although a greater percentage are classified as belonging to scientific newsgroups. Therefore, we fail to find support for our third and fourth hypotheses.

As is seen in Table IV, similar proportions of evolutionary blog documents are assigned to religious and political newsgroups (15.3% and 23.4%, respectively), contrary to expectation. This may be due to the fact that evolutionary blog documents often respond to articles posted on ID blogs, and vice versa, making the subject matter of both sets of blogs interrelated. Although this may be the case, our results still indicate a large proportion of religious and political discussion taking place on blogs that are ostensibly science-focused.

While neither the third nor fourth hypotheses finds support when excluding atheism as a religious category, if one instead chooses to include atheism, the percentage of ID documents classified as either religious or political grows larger than the percentage of documents classified as scientific, increasing to 65.2% (±0.7%). Furthermore, the combined political and religious category percentage for ID blogs grows larger than the corresponding percentage for mainstream science blogs (65.2% verses 60.8%, respectively). Thus, the inclusion or exclusion of atheism from the religious category changes the results significantly.

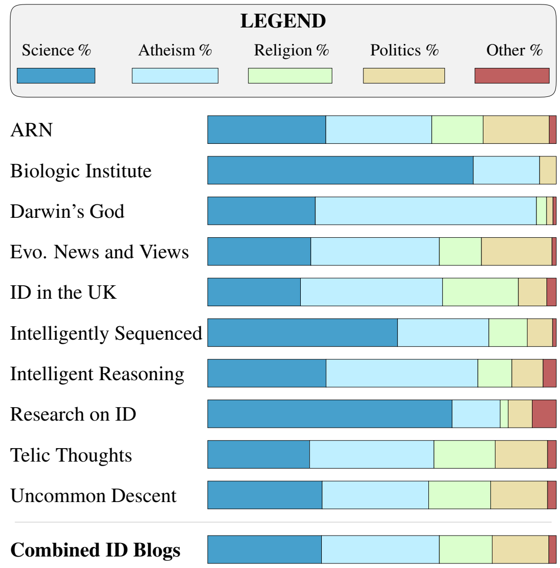

Fig. 2. Classification proportions for ID blog documents.

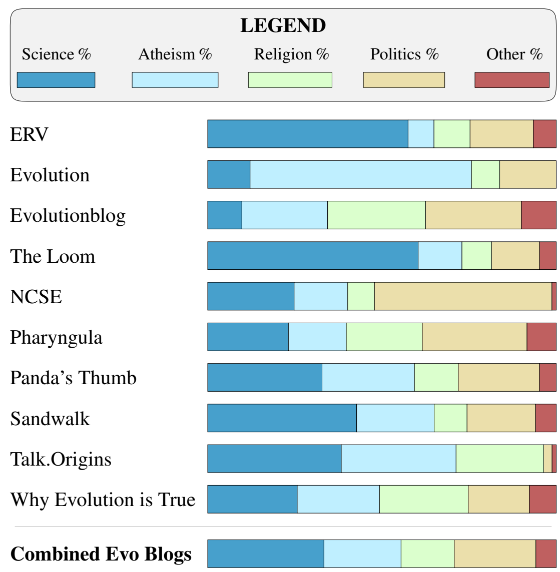

Fig. 3. Classification proportions for evolutionary blog documents.

C. Topic Modeling and Predictive Words

To better understand the distribution of topics within the ID and evolutionary science blog datasets, we trained a naïve Bayes classifier (without LCO features) on the entire 20 newsgroups dataset as well as including two new classes, *ID* and *Evo*, which

TABLE III

CLASSIFICATION GROUP TYPE ASSIGNMENTS FOR ID BLOG DATASET.

Blog	# Docs	% Science	% Religion	% Atheism	% Politics
Uncommon Descent	7,892	32.8	17.8	30.6	16.3
Evolution News and Views	3,546	29.6	12.1	36.9	20.2
Access Research Network	3,066	33.9	14.7	30.4	18.9
Telic Thoughts	2,175	29.2	17.6	35.7	15.0
Darwin's God	651	30.9	2.9	63.4	1.8
ID in the UK	184	26.6	21.7	40.8	8.2
Intelligently Sequenced	481	54.5	11.0	26.2	7.3
Intelligent Reasoning	636	34.0	9.7	43.6	9.0
Research on ID	87	70.1	2.3	13.8	6.9
Biologic Institute Blog	21	76.2	0.0	19.0	4.8
Combined (all ID blogs)	18,739	32.7	15.2	33.8	16.2

TABLE IV

CLASSIFICATION GROUP TYPE ASSIGNMENTS FOR EVOLUTIONARY SCIENCE BLOG DATASET.

Blog	# Docs	% Science	% Religion	% Atheism	% Politics
Panda's Thumb	4,188	32.8	12.6	26.5	23.3
Why Evolution is True	3,029	25.7	25.5	23.6	17.6
Evolutionblog	1,110	9.8	28.1	24.6	27.5
NCSE Blog	1367	24.8	7.7	15.4	50.9
Sandwalk	3,454	42.7	9.4	22.2	19.7
Talk.Origins	167	38.3	25.1	32.9	2.4
ERV	821	57.5	10.4	7.4	18.1
Pharyngula	596	23.2	21.8	16.6	30.0
Evolution	74	12.2	8.1	63.5	16.2
The Loom	669	60.4	8.5	12.6	13.8
Combined (all evolution blogs)	15,475	33.3	15.3	22.1	23.4

Fig. 5. One-hundred most predictive words for the ID and evolutionary datasets, respectively.

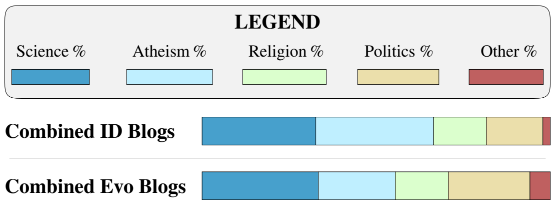

Fig. 4. Comparison between the combined ID and combined evolutionary blog classification proportions.

contained all documents from both blog datasets. We then found the one-hundred most predictive words for ID and evolutionary science documents, ranked by $p(\text{word}|\text{class})/\sum_j p(\text{word}|\text{class}_j)$, where class_j where class_j is taken over all other classes, excluding the class in question. Figure 5 displays the most predictive words for the ID document class (left) and the most predictive words for the evolutionary science blog document set (right), with greater size indicating a more predictive word. Words such as "O'leary" and "Darwinists" are strong predictors of ID documents, while words relating to the Freshwater trial [29] strongly predict evolutionary science blog documents.

We also applied Latent Dirichlet Allocation [30] to the two blog datasets, using the *gensim* [31] package for python. We trained ten topic models for the ID dataset and ten topic models for evolutionary science blog dataset, which are listed in Table V. Automated LDA inference reveals topics dealing with religion (even for the small number of topics trained), such as "God and science" and "Religion and Atheism", and topics dealing with politics and public policy, such as "ID in schools" and "Creationism in schools". Three of the ten main topics from the evolutionary science blog dataset appear to deal with religion, with one topic from the ID dataset also containing religious vocabulary. These qualitative findings support our quantitative classification results by indicating the existence of content dealing with both religion and politics within the document sets.

VI. Discussion

Our results support two of our initial hypotheses, and may provide support for all four, if atheism is classified as a religion. By using a naïve Bayes classifier trained on the 20 newsgroups dataset, we were able to categorize intelligent design blog documents into religious, political and scientific categories, and found a significant proportion of documents (31.4%) assigned to either religious or political categories. When atheism is included as a religion, this percentage increases to 65.2%. Surprisingly, the evolutionary science blog dataset also contains a significant portion of religious and political content, up to 60.8% when including atheism as a religious category.

Although supervised classifiers are often reliable for text categorization when trained with sufficient data, one must exercise caution when drawing strong conclusions from our results. Classification by an automated classifier does not constitute absolute proof of scientific or religious nature, due to the unavoidable presence of type I errors. However, evaluating the accuracy of our methods on datasets where labels are known increases our confidence in these results. As an additional caveat, we must also take into account the interdependent nature of the blogs within our datasets, due to cross-posting among blogs, as well as blog posts written to critique opposing views. Nevertheless, the results are intriguing and strongly suggest a significant presence of religious and political content in science-related blogs dealing with the topic of evolution.

VII. Conclusion

Asking humans to estimate how much religious and political content exists in social media postings runs the risk of revealing more about the leanings of the classifier (and their unconscious prejudices and biases) than anything about the content itself. Our insight is that a supervised classifier built to categorize text and trained on separate religious, political, and scientific datasets, can also be used to categorize contentious data. In doing so, we seek to reduce the risk of assessor bias, as computers are less likely to be persuaded by social pressures. While researcher bias may become unconsciously intertwined in even automated methods [32], the bias present in human judges is unavoidable; thus, our strategy holds promise for reducing the bias of human judgements in such settings, in favor of more principled (and interpretable) classification techniques. By asking clear questions, using standard and well-known machine learning methods (e.g., naïve Bayes classifiers), and evaluating publicly available data sources, this risk becomes reduced. Our results are not to be taken as the final word on this topic, but merely as an interesting example of using simple tools to investigate important sociological questions. More complex analyses can (and should) be performed using these and other datasets, and our work can serve as a template for posing hypotheses amenable to testing by quantitative methods.

References

[1] J. D. Brown, "Evaluations of self and others: Self-enhancement biases in social judgments," *Social cognition*, vol. 4, no. 4, p. 353, 1986.

TABLE V

LDA topics learned from ID and evolutionary science blog datasets. Words are ordered by their likelihood within a topic, with more probable words appearing first. Topic labels are guessed from the words.

ID topics	Most probable words in topic
“Cognitive science”	brain, people, human, life, years, time, psychology, news, mind, ud
“Academia”	science, university, evolution, public, scientific, article, scientists, people, education, research
“Human evolution”	darwin, human, darwinism, people, darwins, selection, evolutionary, man, natural, theory
“God and science”	science, god, evolution, scientific, theory, people, universe, world, nature, evidence
“Intelligent Design”	design, nature, intelligent, designed, life, system, natural, information, intelligence, systems
“ID in schools”	design, intelligent, evolution, science, theory, scientific, evidence, school, students, biology
“Cellular information”	evolution, information, dna, genes, gene, cell, selection, protein, mutations, complex
“Origin of life”	life, earth, origin, evolution, early, evidence, years, evolutionary, rna, cambrian
“Fossil record”	evolution, species, years, evolutionary, fossil, darwin, human, humans, evidence, common
“Climate change”	global, climate, warming, baylor, science, research, change, universe, time, data

Evolutionary blog topics	Most probable words in topic
“Bacteria”	food, bacteria, years, disease, life, water, energy, carbon, plants, molecule
“Molecular biology”	protein, 1, dna, 2, proteins, molecule, rna, sequence, acid, cell
“Science vs. Religion”	science, religion, evolution, religious, scientific, faith, scientists, people, god, public
“Religion and atheism”	god, people, time, good, religion, faith, atheists, world, science, years
“Creationism in schools”	design, science, evolution, intelligent, religious, scientific, school, students, freshwater, creationism
“Academic research”	university, science, nobel, prize, professor, free, time, research, comments, people
“Evolution”	evolution, selection, theory, evolutionary, natural, species, life, science, change, process
“Genetics”	genes, gene, species, genome, dna, time, genetic, fish, human, paper
“God and evolution”	god, design, evolution, evidence, life, natural, argument, universe, world, human
“Evolution of species”	evolution, species, dna, book, human, paper, evolutionary, science, years, evidence

[2] H. Blumer, “Race prejudice as a sense of group position,” in *New Tribalisms*. Springer, 1998, pp. 31–40.

[3] R. J. Chenail, “Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research,” *The Qualitative Report*, vol. 16, no. 1, p. 255, 2011.

[4] F. G. Mármol, M. G. Pérez, and G. M. Pérez, “Reporting offensive content in social networks: toward a reputation-based assessment approach,” *IEEE Internet Computing*, vol. 18, no. 2, pp. 32–40, 2014.

[5] W. Cohen, “Learning rules that classify e-mail,” in *AAAI Spring Symposium on Machine Learning in Information Access*, vol. 18, 1996, p. 25.

[6] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, “A bayesian approach to filtering junk e-mail,” in *Learning for Text Categorization: Papers from the 1998 Workshop*. Madison, Wisconsin: AAAI Technical Report WS-98-05, 1998. [Online]. Available: http://robotics.stanford.edu/users/sahami/papers-dir/spam.ps

[7] N. Jindal and B. Liu, “Review spam detection,” in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 1189–1190.

[8] E. Spertus, “Smokey: Automatic recognition of hostile messages,” in *Proceedings of the National Conference on Artificial Intelligence*. JOHN WILEY & SONS LTD, 1997, pp. 1058–1065.

[9] A. Razavi, D. Inkpen, S. Uritsky, and S. Matwin, “Offensive language detection using multi-level classification,” *Advances in Artificial Intelligence*, pp. 16–27, 2010.

[10] G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, “Detecting offensive tweets via topical feature discovery over a large scale twitter corpus,” in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 1980–1984.

[11] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, “Exploiting burstiness in reviews for review spammer detection.” *ICWSM*, vol. 13, pp. 175–184, 2013.

[12] Y. Zhou and W. M. Inge, “Malware detection using adaptive data compression,” in *Proceedings of the 1st ACM workshop on Workshop on AISec*, ser. AISec ’08. New York, NY, USA: ACM, 2008, pp. 53–60. [Online]. Available: http://doi.acm.org/10.1145/1456377.1456393

[13] D. Chau, C. Nachenberg, J. Willhelm, A. Wright, and C. Faloutsos, “Polonium: Tera-scale graph mining and inference for malware detection,” in *Proccedings of SIAM International Conference on Data Mining (SDM) 2011*, 2011.

[14] A. Mukherjee, B. Liu, J. Wang, N. Glance, and N. Jindal, “Detecting group review spam,” in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 93–94.

[15] T. Lappas, “Fake reviews: The malicious perspective,” in *International Conference on Application of Natural Language to Information Systems*. Springer, 2012, pp. 23–34.

[16] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, “Spotting fake reviews via collective positive-unlabeled learning,” in *2014 IEEE International Conference on Data Mining*. IEEE, 2014, pp. 899–904.

[17] B. Forrest and P. R. Gross, *Creationism’s Trojan Horse*. Oxford University Press, feb 2004. [Online]. Available: http://www.oxfordscholarship.com/view/10.1093/acprof:

oso/9780195157420.001.0001/acprof-9780195157420
[18] M. Behe, *The edge of evolution : the search for the limits of Darwinism*. New York: Free Press, 2008.
[19] S. Meyer, *Signature in the cell : DNA and the evidence for intelligent design*. New York: HarperOne, 2009.
[20] D. Axe, *Undeniable : how biology confirms our intuition that life is designed*. New York, New York: HarperOne, 2016.
[21] T. Joachims, "A probabilistic analysis of the rocchio algorithm with tfidf for text categorization," Carnegie Mellon University, Tech. Rep. CMU-CS-96-118, 1996.
[22] P. Domingos and M. Pazzani, "On the optimality of the simple bayesian classifier under zero-one loss," *Machine learning*, vol. 29, no. 2, pp. 103–130, 1997.
[23] D. Guthrie, B. Allison, W. Liu, L. Guthrie, and Y. Wilks, "A closer look at skip-gram modelling," in *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, 2006, pp. 1–4.
[24] J. Rennie, "The 20 newsgroups data set - reduced (18828) with header information removed," Jan. 2008, http://people.csail.mit.edu/jrennie/20Newsgroups/.
[25] D. Yekutieli and Y. Benjamini, "Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics," *Journal of Statistical Planning and Inference*, vol. 82, no. 1-2, pp. 171–196, 1999.
[26] D. DeWolf, *Traipsing into evolution : intelligent design and the Kitzmiller v. Dover decision*. Seattle, WA: Center for Science Culture, Discovery Institute, 2006.
[27] M. Chapman, *40 days and 40 nights : Darwin, intelligent design, God, Oxycontin, and other oddities on trial in Pennsylvania*. New York, NY: Collins, 2007.
[28] J. Calvert, "Kitzmiller's error: Defining religion exclusively rather and inclusively," *Liberty UL Rev.*, vol. 3, p. 213, 2009.
[29] Wikipedia, "John Freshwater — Wikipedia, the free encyclopedia," http://en.wikipedia.org/w/index.php?title=John%20Freshwater&oldid=755317401, 2017, [Online; accessed 06-January-2017].
[30] D. Blei, "Introduction to probabilistic topic models," *Communications of the ACM*, pp. 1–16, 2011.
[31] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, May 2010, pp. 45–50.
[32] M. Shepperd, D. Bowes, and T. Hall, "Researcher bias: The use of machine learning in software defect prediction," *IEEE Transactions on Software Engineering*, vol. 40, no. 6, pp. 603–616, 2014.