Crowdsourcing to Analyze Belief Systems Underlying Social Issues

J. Hunter Priniski (priniski@ucla.edu)
Department of Psychology
University of California, Los Angeles

Keith J. Holyoak (holyoak@lifesci.ucla.edu)
Department of Psychology
University of California, Los Angeles

Abstract

People's beliefs and attitudes about social and scientific issues, such as capital punishment and climate change, appear to form complex but generally coherent networks. Understanding the nature of these networks is a prerequisite for designing interventions for changing beliefs on the basis of rational arguments and evidence. It is therefore important to develop methods to represent and analyze the form and nature of belief networks, which may not be explicitly verbalizable. Adopting an emerging approach that utilizes crowdsourcing to develop educational interventions, we mined discussions from the Reddit forum Change My View to determine which beliefs and types of information underlie people's attitudes about capital punishment. By combining computational analyses based on a topic model with more qualitative assessments of the extracted topics, we found that moral arguments are more prevalent than statistical or data-based arguments. The present study serves as a test case for the open sourced software `crowdpy`, a Python toolkit for running naturalistic studies on the web, which will enable other researchers to use crowdsourcing in their research. This approach sets the stage for research exploring potential interventions to change people's beliefs.

Keywords: crowdsourcing, digital field studies, belief networks, attitude change

Introduction

The processes by which people form and change their beliefs is central to understanding human thinking. As a consequence of the recent increase in political and scientific misinformation propagated via the internet, it is important to leverage the cognitive science of belief change to develop interventions that can effectively mitigate the resulting social ills and misconceptions, such as xenophobia, racism, and science denial.

The most common approach taken by cognitive scientists aiming to correct misconceptions is to experimentally test the efficacy of different types of corrective information on people's attitudes (e.g., Horne, Powell, Hummel, Holyoak, 2015; Nyhan & Reifler, 2015; Ranney & Clark, 2016). However, people may be resistant to corrective educational information, and educational interventions developed by academic researchers and tested in the lab have often proved ineffective (e.g., Lai et al., 2014). The prevalence of negative results has led some researchers to conclude that changing beliefs based on rational assimilation of evidence is not feasible (e.g., Flynn, Nyhan, & Reifler, 2017).

But before accepting this pessimistic conclusion, it is important to consider the complex nature of people's beliefs. In general, people do not hold individual beliefs in isolation, but rather systems of beliefs that mutually cohere with one another (Thagard, 1989). Because people aim to maintain coherence in their belief systems, changing one belief can potentially alter interrelated beliefs in the overall network (Holyoak & Simon, 1999), including beliefs related to moral issues (Holyoak & Powell, 2016). For many social and scientific topics, such as vaccine safety and climate change, analyzing individual beliefs in relation to a larger ecosystem of related beliefs can shed light on people's conceptual frameworks.

Graphical models are frequently used to formally represent belief systems, with commonly-held beliefs represented as nodes and relationships between beliefs (typically, positive or negative implications) coded as directed edges. Recent work has developed techniques for eliciting the form of people's belief networks (Powell, Weisman, & Markman, 2018). Here we consider the role that crowdsourcing may play in gaining a better understanding of the belief networks that support views about complex social issues. The present study focuses on simply identifying the basic types of beliefs that underlie disagreements about a controversial social issue, capital punishment. Our findings may facilitate future work aimed at either finding effective rational arguments for changing beliefs or fleshing out complete graphical models of people's belief systems surrounding capital punishment.

Crowdsourcing as a Research Tool

Crowdsourcing affords a number of advantages as a tool for cognitive science. One is ecological validity, as materials acquired by crowdsourcing by their very nature are vetted naturalistically (cf. Kahan & Carpenter, 2017). Furthermore, crowdsourcing allows researchers to test the ecological validity of laboratory findings.

In addition, crowdsourcing can take advantage of free and plentiful big data. Huge numbers of people discuss social issues online, and crowdsourcing enables researchers to mine this information to assess what arguments and information are effective at changing people's beliefs, as well as to identify beliefs most central to common misconceptions.

Crowdsourcing with `crowdpy` We ran the present crowdsourcing study using the open source software package `crowdpy`: a flexible and easy-to-use Python toolkit designed for running crowdsourcing and digital field studies (Priniski, 2020). Despite an emerging interest in the ecological validation of lab studies (Kahan & Carpenter, 2017), crowdsourced studies remain relatively rare in the cognitive science literature. This mismatch may stem from naturalistic research being difficult to run—due to a required familiarity with data mining and machine learning—and being analytically flexible, which in turn casts doubt not only on the replicability of such studies but their value to science

more broadly. `Crowdpy` is designed to help overcome these shortcomings by providing researchers with an intuitive toolkit for running and preregistering naturalistic studies. Specifically, `crowdpy` provides support for (1) specifying and executing the computations underlying a crowdsourcing study, and (2) easy access to common data mining tools (e.g., accessing data from Reddit and Twitter). We hope the present paper will demonstrate the value `crowdpy` can provide as part of the cognitive scientist's toolkit. For more information and tutorials on using `crowdpy`, the reader should visit the software's website at crowdpy.com.

Crowdsourcing to Analyze Belief Networks

Here we describe a general methodology that can enable researchers to crowdsource the content and nature of belief systems using the `crowdpy` toolkit. Specifically, this methodology will allow researchers to (1) discern the reasons, evidence, and auxiliary beliefs on which people's views about controversial issues depend, and (2) determine which types of auxiliary beliefs are most central, and therefore perhaps most important as targets for counterarguments. To aid this goal we have created a ready-to-go `crowdpy` workflow designed to help researchers with little-to-no coding experience use crowdsourcing to analyze belief networks, which can be found on the `crowdpy` website.

In broad strokes, the method of crowdsourcing flexible beliefs follows two steps:

1. Mine social media (here, we focus on the Reddit forum Change My View, described below)
2. Train an unsupervised topic model (e.g., Latent Dirichlet Allocation; Blei et al., 2003) to determine which beliefs and reasons underlie people's core beliefs.

Additional data analyses or follow-up behavioral studies can then be performed to assess which beliefs are most flexible and which types of arguments are most effective in changing them. We will now elaborate on the above two core steps, and describe how they are executed in the crowdsourcing software tool created for this task.

Step 1: Mine Naturalistic Data An approach commonly used to mine social media data is to access a social media website's Application Programming Interface (API). APIs provide direct access to the platform's freely available data. The crowdsourcing software tool we developed consists of the code necessary to access the Reddit and Twitter APIs, so that a user without prior data-mining experience can mine these platforms. To access the APIs with our crowdsourcing software tool, the user will only have to create an account with the respective platform and specify which call they want to make (e.g., specify a keyword they would like to search for in the data). The resulting data will be saved automatically in a csv or excel file.

Step 2: Run a Topic Model A topic model serves to analyze which topics are frequently discussed for a given issue (e.g., climate change, capital punishment, vaccine safety), making it possible to determine the auxiliary beliefs and reasons on which people's attitudes towards complex social and scientific issues depend. These topics will serve as the basis for assessing which auxiliary beliefs support people's representations of a complex social or scientific issue. In conjunction with the code that connects users to the website's APIs, the crowdsourcing software tool will perform topic modeling using LDA (Blei et al., 2003). To assist in interpreting the model, the software tool will also create an interactive data visualization using the Python package pyLDAvis (Maybe, 2019), which allows users to examine the topic details in greater clarity.

Researchers can crowdsource from numerous social media platforms. Here we focus on analyzing data from the Reddit forum Change My View. Given the structure and nature of conversations on Change My View, it is at present the most natural data source to use for our present purposes.

Reddit's Change My View

Change My View is a popular Reddit forum in which users post their views on issues ranging from gun control to the movies. "Redditors" posting in this community understand that others will attempt to change their view by providing arguments opposing their beliefs. Because some arguments are more persuasive than others, the variance in argument quality found on the forum provides a naturalistic resource for analyzing the features of effective arguments (e.g., Priniski & Horne, 2018; Hidey & McKeown, 2018). Work in computer science has focused on automatic extraction of features that predict the probability an argument will be effective (Tan, Niculae, Danescu-Niculescu-Mizil, & Lee, 2016), and on identifying aspects of beliefs that are most amenable to change (Jo et. al, 2018).

Previous work has demonstrated that researchers can build on successful crowdsourced arguments—that is, arguments mined from naturalistic resources such as Reddit, Facebook, and Twitter—to develop effective educational interventions likely to correct people's misconceptions (e.g., Priniski & Horne, 2019). Narrowing the hypothesis space of possible interventions increases the probability that researchers will be able to develop an effective intervention. Indeed, Priniski and Horne (2019) demonstrated that crowdsourcing can produce educational interventions just as effective at correcting misconceptions as interventions published in top academic journals.

The present study was designed to demonstrate how the crowdsourcing approach can be used to identify the types of beliefs and reasons that underlie people's attitudes about capital punishment. The goal is to better understand how people's belief systems are structured, and how core beliefs cohere with additional auxiliary beliefs.

Table 1: *Summary of Change My View data used in study.*

Position	Number of posts	Number of comments
Pro-capital punishment	94	5264
Anti-capital punishment	55	3921
Non-related	32	2111
Total	**181**	**11296**

Crowdsourcing Study: Assessing Beliefs About Capital Punishment

The present study aimed to shed light on the structure of people's belief systems surrounding capital punishment, and what types of reasons they deem relevant to their attitudes. Understanding the beliefs and justifications that underlie people's attitudes toward capital punishment is a prerequisite for developing effective interventions that might change their attitudes.

In broad strokes, we first mined Reddit for Change My View discussions relating to capital punishment. We then hand-labeled the discussions as either promoting an attitude in favor of capital punishment or an attitude against it. Second, we employed a topic modeling algorithm, LDA, to elucidate the beliefs and justifications on which people's attitudes to capital punishment hinge. Third, we tested a hypothesis inspired by the results obtained using the topic model. We assessed whether people's attitudes towards capital punishment (on both sides) rely more strongly on morality-based justifications, in contrast to data-driven or statistical justifications.

Data Collection

The first step in our crowdsourcing methodology is to mine naturalistic data. We mined discussions about capital punishment from the Reddit forum Change My View. This forum provides a straightforward platform for obtaining data that can be used to analyze attitudes, attitude change, and persuasion tactics in a relatively naturalistic setting (Priniski & Horne, 2018).

To this end, we mined Change My View posts related to capital punishment and the death penalty using the Python Reddit API Wrapper (2019). Specifically, we searched Change My View for discussion posts and replies for the strings "capital punishment" and "death penalty". If either the post or a reply returned a match, we collected the full discussion for our dataset. Descriptive statistics for the mined dataset are presented in Table 1 above.

Topic Modeling

Method We fit an unsupervised topic modeling algorithm, LDA, to determine which types of beliefs and considerations are most relevant to attitudes toward capital punishment expressed in the discussion posts. LDA assigns documents (here, the titles of discussion posts on Change My View) to topics by instantiating a set of k topics, where each topic is defined by a set of words that compose the documents. In general terms, the algorithm assigns a probability value for how likely a document belongs to each topic. This probability value is determined by how many of the topic's "representative" words appear in the document.

Table 2: *Example post in Change My View dataset.*

Title (i.e., attitude)	Selftext (i.e., justification)
CMV: The death penalty is only a harsh punishment for people who are wrongly convicted. For the guilty, it is by no means the ultimate punishment. It is inherently unjust and should be universally abolished.	This is USA-specific, but feel free to include your perspective if you are not a USA-ian. If someone killed or hurt someone in the circle of people I consider "family", I would probably want . . .

Because divergent attitudes toward capital punishment may hinge on different beliefs, reasons, or justifications, we fit one model to the titles of posts indicating a favorable attitude towards capital punishment, and another model to the titles of posts indicating disapproving attitudes, using the Python package `Genisim` (Řehůřek & Sojka, 2010). We fit the model to the titles because these provide a concise, non-noisy representation of the participant's attitude and justifications. The topic model could also be fit to the `selftext` of the posts (a more detailed justification for their attitudes, written by the posting user). An example title on which we trained the topic model is provided in Table 2.

Results We fit two topic models, one to the pro-capital punishment posts and another to the anti-capital punishment posts, each with five topics. The topics underlying the pro-capital punishment attitudes are displayed in Table 3 (the topics from the anti-capital punishment discussions are nearly identical). Model selection was guided by maximizing a model's coherence scores (Řehůřek & Sojka, 2010), which numerically quantify on a scale from 0 to 1 how well the topics fit to the data and how well the terms within a single topic are semantically related. The coherence value of the selected model for pro-capital punishment attitudes was .48; coherence for anti-capital punishment attitudes was .44. In addition to using a model's coherence scores, model selection was also guided by a subjective assessment of the interpretability and meaningfulness of the topic classifications, guided by previous analysis of the considerations many people deem relevant to their stance on capital punishment (e.g., Miske et al., 2019). The selected topic model can be found at the project's GitHub repository.

The topics (and their associated keywords) found to underlie both pro- and anti-capital punishment attitudes indicate that people more strongly consider ethical and moral justifications than statistical and cost-benefit reasons. Both types of justifications are factors commonly identified in work on psychology and law as underlying people's attitudes toward capital punishment (e.g., Miske et al.,

Table 3: *Words associated with pro-capital punishment attitudes.*

	Top 10 keywords
Topic 0	life, prison, give, crime, sentence, **wrong**, form, abolish, **morally**
Topic 1	crime, prison, *child*, sentence, life, make, favor, oppose, *rape*, standard
Topic 2	**deserve**, life, believe, murder, pro, crime, state, citizen, criminal, people
Topic 3	legal, people, murder, state, system, time, criminal, believe, **innocent**, kill
Topic 4	believe, use, method, crime, must, viable, consider, certain, state, replace

Note. **Bolded** words indicate a moral basis for capital punishment, whereas *italicized* words indicate the role capital punishment plays in achieving retribution. Retributive and other morally-grounded beliefs appear to underlie each topic. Words associated with anti-capital punishment attitudes are similar to those associated with pro-capital punishment attitudes and also emphasize moral coherence over statistical concerns.

2019). The observed emphasis on moral and ethical justifications (reflected in keywords such as "morally", "barbaric", and "wrong") over statistical and cost-benefit justifications (keywords such as "cost", "rate", or "crime statistics") suggests that coherence of moral beliefs plays a central role in supporting people's attitudes towards capital punishment (cf. Holyoak & Powell, 2016). In the next section we report a more direct test of this hypothesis.

Statistical Analyses

To assess whether attitudes toward capital punishment depend on moral coherence to a greater extent than statistical evidence, we adapted an approach developed by Priniski and Horne (2018) to measure evidence-related and statistical language in Change My View discussions. Our aim was to determine whether participants used language more indicative of reasoning driven by moral coherence as compared to statistical evidence. We used two dictionaries of terms commonly associated with types of reasoning to measure types of reasoning in the posts. To measure rates of moral coherence, we used the 2nd version of the Moral Foundations Dictionary (Frimer, Boghrati, Haidt, Graham, & Dehgani, 2019).

To our knowledge no comparable inventory of terms associated with everyday statistical reasoning exists. We therefore employed a data-driven approach (extracting features from a Random Forest model) to create a list of terms indicative of statistical reasoning. To this end, we first created a dataset composed of Reuters financial news articles (Ding et al., 2014), a large set of discussions from the Reddit forum r/Statistics, and tweets collected by Go, Bhayani, and Huang (2009) that have commonly been used to build sentiment analysis models. The Reuters and r/Statistics corpora involve reasoning about numeric values and statistics, while the tweet corpus provides a reference group against which to classify the Reuters r/Statistics posts. With this grouped dataset ready, we then trained a Random Forest classifier to determine which dataset (Reuters or r/Statistics vs. Twitter) each post in the conjoined dataset came from, based on the text's individual word tokens. Having created a model with high predictive accuracy, we then extracted the features that the model used to determine which posts were in the Reuters or r/Statistics datasets and used that as the basis for our dictionary of terms related to statistical reasoning. The code used to mine the Reddit forum, prepare and model the data using Random Forest classification can be found at this paper's GitHub repository (linked above). We sketch the statistical details of this approach to feature selection below.

Feature Extraction with Random Forests

Random Forests are a class of supervised, ensemble learning algorithms that are commonly used in both classification and regression tasks. They model data by aggregating the results of a large set of decision trees constructed from random subsets of the feature space. In addition to their widespread success in machine learning applications, they are also an effective variable selection tool. Here, we sought to use Random Forests to highlight which terms are most strongly associated with statistical reasoning. To achieve this goal, we first fit a Random Forest classifier to the compiled dataset described above. Next, we extracted the features the model identified as the most important to determining classifications.

Here, we sketch the statistical framework underlying this process and relate it to the task at hand. Let $\mathcal{D} = \{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_n, y_n)\}$ be the collection of document tuples $d_i = (\boldsymbol{x}_i, y_i)$ in our corpus (i.e., the conjoined dataset). Each document d_i is composed of an encoding of the text data, $\boldsymbol{x}_i$, and its associated classification, $y_i \in \{0, 1\}$ ($y_i = 0$ indicates d_i is from the Twitter dataset while $y_i = 1$ indicates the document belongs to either the Reuters or r/Statistics datasets). Here, $\boldsymbol{x_i}$ is a bag-of-words of document d_i.

The principle idea behind a Random Forest classifier is to aggregate the results of many smaller *classification trees* to derive predictions. In their simplest form, classifications trees are binary decision trees, which for each node in the tree, must make a binary decision of whether a document belongs to a class (i.e., document indicates statistical reasoning) or not. Nodes in a classification tree consists of a random subset of features or dimensions from $\boldsymbol{x}$, and classification decisions are made at each node in the tree.

The importance of each feature loosely stems from how essential that feature is to maximizing the model's classification accuracy. Roughly, importance of a feature $x_j \in \boldsymbol{x}$ is assessed by how much the *Gini impurity* is decreased when the data is split at that feature relative to other features. Gini impurity, which is an information theoretic measure that calculates the probability a randomly selected document is misclassified given a set of features, takes the following form:

$$G = \sum_{i=1}^{C} p_i(1 - p_i)$$

with C representing the number of classes. By iterating over a random subset of features, each classification tree seeks to minimize G in order to classify a document. A Random Forest model will then aggregate all of the tree-level classifications for a document and form a final classification based on the class which has the most "votes".

We trained a Random Forest model with the purpose of distinguishing between text documents that likely demonstrate statistical reasoning (i.e., financial news and discussions from the Reddit forum r/statistics) and those that do not (a random set of tweets commonly used for training sentiment analysis models). The corpus was constructed such that any Random Forest model could easily distinguish between both types of documents. Indeed, even base models performed achieved accuracy scores above 95% without any parameter tuning. The features the Random Forest model relied most strongly on to classify the documents, along with their quantitative importance scores learned by the model, can be found on the project's GitHub (linked above). However, because our general goal is to construct a dictionary of terms that map onto statistical reasoning, we refined the returned set of features by (1) removing terms that are artifacts of the dataset and do not reflect statistical reasoning (e.g., "New York" and "Reuters") and (2) adding synonyms and related terms to the remaining set of terms. The revised set of terms, which was used in the following analyses, can be found on the project's GitHub as well.

Predicting Coherence-based and Statistical Considerations

To test our hypothesis that attitudes toward capital punishment depend more on moral coherence than on statistical evidence, we performed Bayesian mixed effects modeling using the `R` package `brms` (Bürkner, 2018). We examined how participants justified their capital punishment attitudes by measuring the rates at which their language pertained to moral-based considerations versus statistics-related considerations. Specifically, we treated each word in the dataset of attitude justifications as belonging to either moral terms (e.g., "morally", "logically", and "wrong"), statistical terms (e.g., "data", "rate", "cost"), or as a stopword (i.e., a term that belongs to neither category). By modeling the rates at which participants justify their beliefs using moral-based or statistics-based language, we can estimate the extent to which moral-based and statistics-based reasons (and beliefs about those reasons) are relevant to people's attitudes toward capital punishment.

As shown in Figure 1, when Redditors justify their beliefs, they use language more strongly associated with moral-based justifications than with statistical justifications (e.g., costs associated with capital punishment or its efficacy at reducing crime rates). This result converges with our findings from the LDA model, indicating that people's attitudes toward capital punishment cohere with their moral considerations.

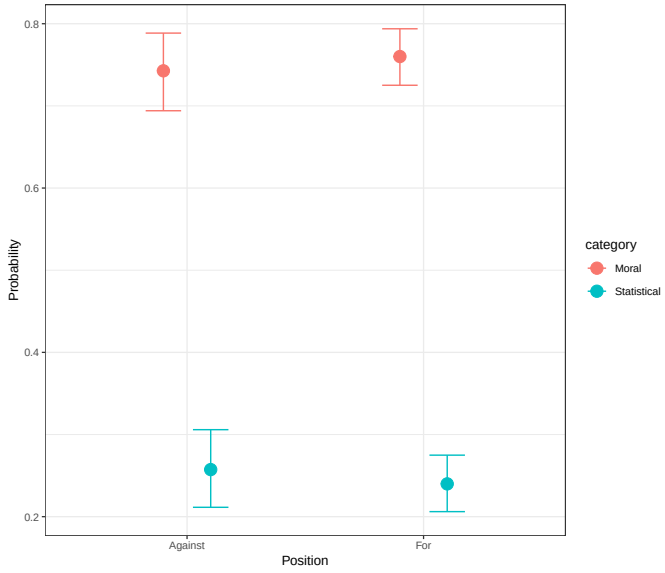

Figure 1: Prevalence of moral and statistical language in posts that either demonstrate pro-capital punishment attitudes or anti-capital punishment attitudes. Error bars represent 95% Bayesian credible intervals.

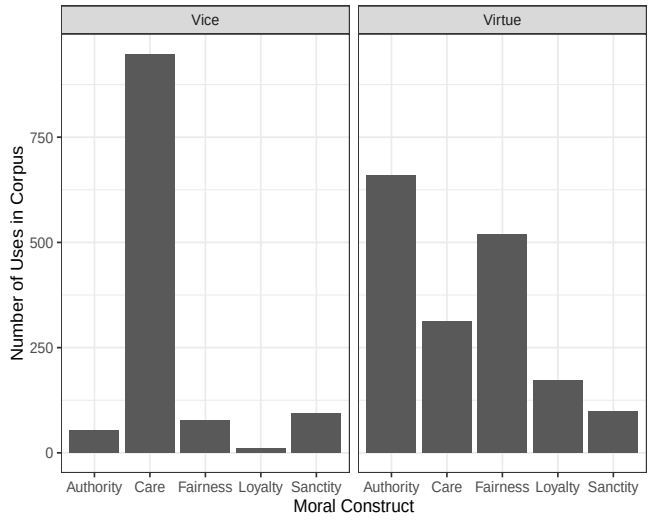

Figure 2: Prevalence of different moral constructs in our corpus. Consideration of *care* (reflecting on the pain of the victims), *authority*, and *fairness* are the most common in the dataset, indicating that people consider these moral values most strongly when justifying their capital punishment positions.

We looked a bit closer at which types of moral constructs are at play in people's moral coherence reasoning. In Figure 2, we see that considerations related to *care* as a vice (i.e., considerations about the suffering and pain inflicted on

victims) is the most prevalent moral consideration. However, the virtues of *authority* and *fairness*, which reflect considerations about respecting order or authority and the importance of achieving justice, respectively, are also strongly considered. Interventions should target these types of beliefs, because they appear to be most central to people's capital punishment attitudes.

Discussion

The beliefs people hold, particularly those relating to moral and scientific issues such as capital punishment and climate change, are not held in isolation. Rather, people's beliefs are situated in an interlocking and cohering system (Thagard, 1989). Here, we propose a crowdsourcing methodology designed to identify the types of beliefs underlying people's representations of moral or scientific topics, focusing on the case of people's attitudes about capital punishment. Specifically, we fit an unsupervised topic model (Latent Dirichlet Allocation) to a dataset of people's attitudes about capital punishment. The results obtained using the topic model suggested that people's attitudes toward capital punishment cohere with their moral considerations to a greater extent than with statistical or data-oriented evidence. We were able to confirm this finding in a follow-up analysis using a different approach based on natural language processing. The results of this study demonstrate the value of using crowdsourcing as the basis for an automated technique to uncover the beliefs and considerations underlying people's attitudes about a controversial social issue.

To facilitate use of our proposed method, we built an easy-to-use crowdsourcing software tool `crowdpy`. This tool will enable researchers with little-to-no coding experience to utilize crowdsourcing in their own research. Using this tool, researchers will be able to easily connect to social media APIs (e.g., Reddit and Twitter), thereby allowing them to mine social media data. Furthermore, `crowdpy` will automatically fit a topic model to the crowdsourced dataset and create an interactive data visualization in pyLDAviz, allowing the researcher to gain a functional understanding of the topical structure of their crowdsourced data.

Limitations and Future Directions

One concern with any crowdsourcing methodology is that the data mined from individuals posting and debating their beliefs on social media may not be representative of the eventual target population. Criticisms of the crowdsourcing approach to developing psychological materials have generally centered around potential population differences between where the data is being acquired (e.g., Change My View) and where it is being applied (e.g., the general population). For certain tasks, such as crowdsourcing educational interventions, this critique has been answered by demonstrating the generalizability of crowdsourced materials in a controlled experiment (Priniski & Horne, 2019). It remains an empirical question whether the beliefs determined by crowdsourcing to underlie people's attitudes toward capital punishment also underlie those of the general population. A simple behavioral study that tests the generalizability of the results gleaned from the analyses presented here is a natural next step for this research.

An additional line of future research will need to articulate the structure of people's systems of beliefs with greater specificity, and in particular, to identify the most central nodes in an individual's system of beliefs. Analyzing individual beliefs in relation to a larger ecosystem of related beliefs can shed light on people's conceptual frameworks and decision-making processes, as well as help develop interventions for correcting misconceptions. This process is challenging, however, both because of the often-complex nature of people's belief systems, and because we lack a strong theory of how beliefs change within a larger conceptual system. Crowdsourcing can assist with the first difficulty, but modeling work will be needed to address the second (cf Powell et al., 2018).

How might crowdsourcing be extended to extract a more complete picture of a belief system? The present findings using the topic model suggest that people commonly justify their attitude toward capital punishment by ethical beliefs (e.g., whether or not it is humane) and the value of its retributive impact. Both of these constructs ("ethical footing" and "retribution") can act as separate nodes in a system of beliefs. A deeper understanding of people's belief systems (i.e., accounting for a larger set of auxiliary beliefs) might be achieved in part by mining more data from other sources (e.g., other Reddit forums, Twitter, opinion sections of newspapers).

Once a more detailed representation of people's belief systems has been obtained, a further goal will be to test the efficacy of different interventions targeting individual beliefs in the crowdsourced network. This process will attempt to identify which of these beliefs are especially malleable (i.e., amenable to change by counterarguments) and also pivotal (i.e., those for which a change would maximally impact the rest of the network). The expectation will be that revision of a pivotal belief will trigger a coherence shift (i.e., updating of other beliefs in the network so as to restore coherence with the one that has shifted). Leveraging networks of belief in the process of developing educational interventions may increase the probability that these interventions will be effective in changing misconceptions.

Belief networks derived by crowdsourcing necessarily are based on data aggregated across a population of users. But for many moral and scientific issues, individual differences in network structure are likely to influence which beliefs are especially malleable and/or pivotal. Future work will need to also find ways to enable crowdsourcing to leverage the power of big data while also accounting for individual differences (e.g., Kwon, Priniski, & Chanda, 2018).

A wide range of social and scientific misconceptions, including climate change denial, racism, and xenophobia, pose serious threats to democracy. In this context, methods that enable crowdsourcing of belief systems can provide an

ecologically valid yet rigorous starting point to aid scientists, educators, and policymakers in developing effective interventions.

Acknowledgments

This research was supported by NSF Grant BCS-1827374.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411.

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Advances in Political Psychology*, *38*, 127–150.

Frimer, J., Boghrati, R., Haidt, J., & Dehgani, M. (2019). *Moral foundations dictionary for linguistic analyses 2.0.* Unpublished manuscript.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12).

Hidey, C. T., & McKeown, K. (2018). Persuasive influence detection: The role of argument sequencing. In *Thirty-Second AAAI conference on Artificial Intelligence.* Menlo Park, CA: AAAI.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, *142*(11), 1179 – 1203.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3 – 31.

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences, USA*, *112*, 10321–10324.

Jo, Y., Poddar, S., Jeon, B., Shen, Q., Rose, C., & Neubig, G. (2018). Attentive interaction model: Modeling changes in view in argumentation. In M. Walker (Ed.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 103 – 116). Stroudsburg, PA: Association for Computational Linguistics.

Kahan, D. M., & Carpenter, K. (2017). Out of the lab and into the field. *Nature Climate Change*, *7*(5), 309 – 311.

Kwon, K. H., Priniski, J. H., & Chadha, M. (2018). Disentangling user samples: A supervised machine learning approach to proxy-population mismatch in twitter research. *Communication Methods and Measures*, *12*(2), 216–237.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E., Joy-Gaba, J. A., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765-1785.

Mansour, R., Hady, M. F. A., Hosam, E., Amr, H., & Ashour, A. (2015). Feature selection for twitter sentiment analysis: An experimental study. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 92–103).

Maybe, B. (2019). *pyLDAvis.* Retrieved from https://github.com/bmabey/pyLDAvis

Miske, O., Schweitzer, N., & Horne, Z. (2019). What information shapes and shifts people's attitudes about capital punishment? In A. Goel, C. Seifert, & C. Freska (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*, 459 – 464.

Powell, D., Weisman, K., & Markman, E. M. (2018). Articulating lay theories through graphical models: A study of beliefs surrounding vaccination decisions. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Priniski, J. H. (2020). *Crowdpy: A Python toolkit for running naturalistic studies on the web.* Retrieved from https://github.com/jpriniski/crowdpy

Priniski, J. H., & Horne, Z. (2018). Attitude change on Reddit's Change My View. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2276 – 2281). Austin, TX: Cognitive Science Society.

Priniski, J. H., & Horne, Z. (2019). Crowdsourcing effective educational intervention. In A. Goel, C. Seifert, & C. Freska (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Python Reddit API Wrapper. (2019). Retrieved from https://github.com/praw-dev/praw

Ranney, M. A., & Clark, D. (2016). Climate change conceptual change: Scientific information can transform attitudes. *Topics in Cognitive Science*, *8*(1), 49–75.

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.*

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In J. Bourdeau, J. A. Hendler, & R. N. Nkambou (Eds.), *Proceedings of the 25th International Conference on World Wide Web* (pp. 613 – 624). New York: Association for Computing Machinery.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467.