Exploring media bias with semantic analysis tools: validation of the Contrast Analysis of Semantic Similarity (CASS)

**Nicholas S. Holtzman · John Paul Schott ·
Michael N. Jones · David A. Balota · Tal Yarkoni**

Abstract Text-analytic methods have become increasingly popular in cognitive science for understanding differences in semantic structure between documents. However, such methods have not been widely used in other disciplines. With the aim of disseminating these approaches, we introduce a text-analytic technique (Contrast Analysis of Semantic Similarity, CASS, www.casstools.org), based on the BEAGLE semantic space model (Jones & Mewhort, *Psychological Review*, 114, 1–37, 2007) and add new features to test between-corpora differences in semantic associations (e.g., the association between democrat and good, compared to democrat and bad). By analyzing television transcripts from cable news from a 12-month period, we reveal significant differences in political bias between television channels (liberal to conservative: MSNBC, CNN, FoxNews) and find expected differences between newscasters (Colmes, Hannity). Compared to existing measures of media bias, our measure has higher reliability. CASS can be used to investigate semantic structure when exploring any topic (e.g., self-esteem or stereotyping) that affords a large text-based database.

Keywords CASS · Individual differences · Language use · Media bias · Semantic space models

N. S. Holtzman · J. P. Schott · D. A. Balota
Washington University in St. Louis,
St. Louis, MO, USA

M. N. Jones
Indiana University,
Bloomington, IN, USA

T. Yarkoni
University of Colorado,
Boulder, CO, USA

N. S. Holtzman (✉)
Department of Psychology, C.B. 1125,
Washington University in St. Louis,
One Brookings Drive,
St. Louis, MO 63130, USA
e-mail: nick.holtzman@gmail.com

Introduction

Quantitative text analysis tools are widely used in cognitive science to explore associations among semantic concepts (Jones & Mewhort, 2007; Landauer & Dumais, 1997; Lund & Burgess, 1996). Models of semantic information have enabled researchers to readily determine associations among words in text, based on co-occurrence frequencies. For example, the semantic association between "computer" and "data" is high, whereas the association between "computer" and "broccoli" is low (Jones & Mewhort, 2007). Such methods allow meaning to emerge from co-occurrences. Despite their frequent successes within cognitive science, rarely have such methods been used in other disciplines (see however, work by Chung & Pennebaker, 2008; Klebanov, Diermeier, & Beigman, 2008; Pennebaker & Chung, 2008). The limited use of these models in related disciplines is partially due to the lack of software that allows efficient comparisons of associations that semantic models provide (e.g., to compare the magnitude of the association between "computer" and "data" to that of "computer" and "broccoli"). Moreover, there is no software that allows researchers to make such comparisons when using one's own corpora. Such practical software could help expedite the research process as well as open up new research avenues to be used in parallel with extant modes of lexical analysis, such as word-frequency analysis (Pennebaker, Francis, & Booth, 2001).

Here we offer such software, which we call Contrast Analysis of Semantic Similarity (CASS). By using the output of previous programs such as BEAGLE (Jones & Mewhort, 2007), which provides concept-association magnitudes, CASS allows a researcher to do two things: (i) compare associations within a model (e.g., the degree to which liberalism is perceived as a positive versus a negative ideology) and (ii) compare associations across groups (e.g., media channels) or individuals (e.g., individual newscasters). Meanwhile it allows researchers to control for baseline information in corpora (e.g., the degree to which conservatism is associated with positive versus negative concepts).

To make it easy to apply our approach to a wider range of topics—attitudes, stereotypes, identity claims, self-concept, or self-esteem—we provide and describe tools that allow researchers to run their own analyses of this type (www.casstools.org). The CASS tools software implements a variant of the BEAGLE model (Jones & Mewhort, 2007), which extracts semantic information from large amounts of text and organizes the words into a spatial map, hence the name "semantic space model." As the model acquires semantic information from text, it begins to provide stable estimates of the relatedness between words and concepts. Since such analyses are agnostic about content, they can be flexibly applied to virtually any research question that can be operationalized as differential associations between words within documents.

Here, we validate the CASS method by exploring the issue of media bias. There is consensus regarding which channels are *relatively* more politically liberal or conservative. For example, most people agree that MSNBC is relatively more liberal than FoxNews. The question is whether CASS can recover the agreed-upon rank-ordering of biases.

It is important to discuss what "bias" means in the CASS method and to describe how CASS effects should be interpreted. From a journalistic point of view, bias is any deviation from objective reporting. Determining whether the reports are objective is not part of the capabilities of the CASS approach. As a result, a zero value CASS bias measure is not necessarily the same thing as objective reporting. Instead, the value zero indicates that a source equally associates conservative and liberal terms with good and bad terms (i.e., no preferential concept association). There are some limitations of this approach. For example, objective reporting may properly discuss a series of scandals involving one political party. According to the rules of journalism, this would not indicate a reporting bias. However, the association of the political party with the negative concepts would lead CASS to reveal a bias against that political party. Thus, it is possible that CASS bias values different from zero can be derived from purely objective reporting. The problem of defining a no-bias point is exceedingly difficult and other approaches to media

bias measurement have similar difficulties (Groseclose & Milyo, 2005).

Unlike absolute CASS effects, relative CASS effects can be interpreted in a straightforward way. If one source gives a positive CASS bias value and another source gives a negative CASS bias value, then a person can properly argue that the first source is biased differently than the second source. (One still cannot identify which source is biased away from objective reporting.) Therefore, although absolute CASS effects contain a degree of ambiguity, comparisons of CASS effects among different sources have a clear meaning about relative biases. Accordingly, we focus primarily on the comparisons.

Given that this is the first study to use the CASS method, we set out to validate the approach. We present three types of analyses: (a) estimating group differences between channels, (b) examining individual differences in bias between newscasters, and (c) exploring the value of our CASS approach in comparison to another popular measure of media bias. To explore group differences, we examine three major cable news channels in the United States: CNN, FoxNews, and MSNBC. These channels are good sources of data for our study because they broadcasted large amounts of political content in a year (2008) that was politically charged—a year during which a presidential election took place. To illustrate the additional utility of our methods for exploring *individual* differences, we also explore semantic spaces for FoxNews host Sean Hannity and his putatively liberal news partner, Alan Colmes.

Given that CASS is designed to uncover differential concept associations, and that media bias should theoretically involve such differences, we expected CASS would reveal the established differences in biases across media channels (Groseclose & Milyo, 2005). Specifically, we hypothesized that MSNBC would be the most liberal, followed by CNN, and then FoxNews; the failure to reveal this rank-ordering would indicate that the method is likely invalid. Secondly, we predicted that Colmes would be relatively more liberal than Hannity—providing another validation check. In the final analysis, we explore how our methods compared with existing measures of media bias (e.g., on reliability).

Contrast Analysis of Semantic Similarity (CASS)

Our CASS technique is based on a difference of differences in semantic space—a logic that has been used widely in the social cognition literature. The general approach is perhaps most known in social cognition due to the success of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), where differences in reaction times are compared under conditions with various pairings of words or concepts. Although the IAT is not perfectly analogous to our methods (e.g., our methods are not necessarily tapping

into implicit processes), the IAT is similar in some important ways. Like the IAT, CASS requires the following components: a representation of semantic space, a set of target words from that semantic space, and an equation that captures associations among targets. For demonstrations of the IAT, see https://implicit.harvard.edu/implicit/ (Nosek, Banaji, & Greenwald, 2006).

In the IAT, experimental stimuli that the participant is supposed to categorize are presented in the middle of a computer screen (e.g., "traditional", "flowers", "welfare", "trash"). These words would be used to represent four super-ordinate categories located at the top of the computer screen (on the left: "republican" and "good"; on the right: "democrat" and "bad"). Participants are asked to sort the stimuli to the appropriate super-ordinate category on the left or right. The speed (and accuracy) with which the presented stimuli are sorted tends to indicate in our example the degree to which one associates republicans with good concepts, and associates democrats with bad concepts. These associations are also derived from the counterbalanced phase when the super-ordinate category labels switch sides: "democrat" is paired with "good" while "republican" is paired with "bad". The main substance of the equation used to compare differential associations involves comparing response times (RTs):

RTs (republican, good; democrat, bad) –
RTs (democrat, good; republican, bad)

In this example, negative output means that a person responds more rapidly to the first configuration, implying relative favoritism for republicans (responses to "republican" and "good" as well as "democrat" and "bad" are facilitated); it is inferred that the participant believes republicans are good and democrats are bad. Positive output means that a person responds more rapidly to the second configuration, implying relative favoritism for democrats. Thus, the IAT (like CASS) provides estimates of individual differences in meaning (whether conservatives are good) through the comparison of concept-associations.

The IAT is not limited to assessing biases in political ideology; it has been used to study many topics. For example, researchers have employed the IAT to explore constructs of interest to social psychologists, such as prejudice (Cunningham, Nezlek, & Banaji, 2004), and it has been used by personality psychologists to study implicit self-perceptions of personality (Back, Schmukle, & Egloff, 2009). In parallel with the development of the Implicit Association Test in the social cognition literature, major advances were occurring in cognitive psychology and semantic modeling that ultimately made it possible to develop CASS.

Estimating the semantic space with BEAGLE

In cognitive science, remarkable progress has been made in the last 15 years towards computational models that efficiently extract semantic representations by observing statistical regularities of word co-occurrence in text corpora (Landauer & Dumais, 1997; Lund & Burgess, 1996). The models create a high-dimensional vector space in which the semantic relationship between words can be computed (e.g., "democrat" and "good"). In the present study, we use such a space to estimate ideological stances.

We estimate the semantic similarity between different concepts using a simplified version of the BEAGLE model described by Jones and Mewhort (2007). In BEAGLE, the distance between two terms in semantic space reflects the degree to which the two terms have correlated patterns of co-occurrence with other words. If two terms tend to occur within sentences containing the same words, then they are considered relatively semantically similar; conversely, if they tend to occur in sentences that use different words, then they are considered relatively semantically dissimilar. Note that terms do not have to co-occur together within the same sentence to be considered semantically similar; high semantic similarity will be obtained if the terms frequently co-occur with the same sets of words. For instance, a given text may contain zero sentences that include both of the words "bad" and "worst", yet if the two words show similar co-occurrence patterns with other words, such as "news" and "sign", then "bad" and "worst" may have a high similarity coefficient.

The semantic similarity between two terms is computed as follows. First, an M x N matrix of co-occurrences is generated, where M is the number of terms for which semantic similarity estimates are desired—the *targets*—and N is an arbitrary set of other words used to estimate semantic similarity—the *context*. For instance, if one wishes to estimate the pairwise semantic similarities between 4 different target terms using a reference set of 4,000 words as the context, then the co-occurrence matrix will have a dimensionality of 4 rows x 4,000 columns. Each cell within the matrix represents the number of sentences containing both the ith target term and the jth context word. For example, if there are 14 instances of the target word "democrat" co-occurring with the context word "Wright" (i.e., Jeremiah Wright, Obama's former pastor and source of much controversy during the 2008 election), then 14 will be the entry at the conjunction of the row for "democrat" and the column for "Wright".

Given this co-occurrence matrix, the proximity between any two target terms in semantic space is simply the correlation between the two corresponding rows in the matrix. We use Pearson's correlation coefficient to index similarity; thus, similarity values can theoretically range from –1.00 to 1.00, where a correlation of zero indicates no

relationship between two words (they are quite far apart in space, e.g., *left–copper*), a correlation closer to one indicates two words that are highly related or synonymous (they are close in space, e.g., *left–democrats, copper–zinc*).

CASS interaction term

An intuitive way to measure media bias would be to simply quantify the strength of association between terms connoting political ideology and those reflecting evaluative judgments. For instance, quantifying the similarity between "republicans" and "good" might indicate the degree of bias in the conservative direction. The key problem is that such an estimate ignores important baseline information; in particular, there is little way of knowing what constitutes a strong or weak association, since it is possible that most words show a positive correlation with "good", and that any positive result is not specific to "republicans". Fortunately, this problem can be corrected by calculating a difference of differences. CASS builds on the semantic space model provided by BEAGLE and offers utilities to calculate the difference of differences. The following equation, which we refer to as a *CASS interaction term*, captures all the relevant baseline information, indicating the differential association of "republicans" with "good" compared to "bad" in contrast to the differential association of "democrats" with "good" compared to "bad":

$$[r(\text{republicans, good}) - r(\text{republicans, bad})] -$$
$$[r(\text{democrats, good}) - r(\text{democrats, bad})]$$

Conceptually, this equation contrasts the degrees of synonymy of the terms. The first set of terms "$r(\text{republicans, good})$" represents the correlation of the vectors for "republicans" and "good"; the second represents that of "republicans" and "bad". The difference between the first two correlations, which make up the first half of the equation, captures bias in favor of the political right; the second half of the equation captures bias in favor of the political left. If the output from the first half is larger, then the text has a conservative bias away from zero, but if the output from the second half is larger, then the text has a liberal bias away from zero.

If the two output values match, then the interaction output is zero, and the corpus contains no bias *as defined by the method*. In this first article employing CASS, it is important to note that we make no grand claim that CASS provides a flawless 0.00 that indicates ideological neutrality—indeed, there are specific problems with such an assertion. It is possible, for example, that the target words could have connotations that are independent of political information. The target political concept "conservative" is not necessarily a purely political word. For example, "conservative" can refer to careful and constrained decision-making, which

carries little direct indication of one's political ideology, yet may carry a valence-laden connotation. ("She is conservative and judicious in her decision-making, which leads to good long-term outcomes.") Indeed, the target words may carry connotations from irrelevant semantic contexts, which would alter the expected valence of the target words. Consequently, the true neutral-point would be away from 0.00 (slightly more positive, given the parenthetic example above). Nevertheless, the fact that semantic analysis involves polar-opposite positive and negative concepts (e.g., good, bad) that happen to allow for the emergence of a neutral 0.00 is enticing for research in need of an impartial arbitrator, such as media bias research.

Implementation

The methods described above are implemented in an open-source software package—CASS Tools—written in the Ruby programming language. We have made the software freely available for download at http://www.casstools.org, where there is extensive documentation for the tools as well as exercises for beginners.

Method

To evaluate the CASS technique and explore media bias, we (a) examined biases in each of three channels, (b) explored biases between two prominent individuals in the media, and (c) compared the CASS approach to one popular available approach (think tank analysis).

Obtaining text files Transcripts for our analysis were downloaded from LexisNexis Academic. We obtained all 2008 television transcripts indexed by the keyword "politics" originally televised by MSNBC, CNN, and FoxNews. CNN transcripts contained 48,174,512 words. FoxNews transcripts contained 9,619,197 words. MSNBC transcripts contained 7,803,311 words. Transcripts for Hannity and Colmes were also downloaded from LexisNexis Academic. All available transcripts were used, from October 6, 1996 to January 9, 2009. The Hannity transcripts contained 4,607,282 words and the Colmes transcripts contained 3,773,165 words. All word counts were calculated prior to the handling of negations (see below). Because the unit of analysis in BEAGLE is the sentence, text files were parsed into sentences; sentences were stripped of punctuation using custom automated software.

Variations of the indices In using CASS, it is crucial to measure underlying concepts in multiple ways—for the

same reason that good surveys usually contain multiple items. Therefore, variations of the target words were employed for the interaction terms. Because word frequency has a significant impact on the stability of the effects, we chose to use the most frequently occurring synonyms of the target words: good (great, best, strong); bad (worst, negative, wrong); republicans (conservatives, conservative, republican); democrats (left, democratic, democrat). All possible combinations of these 16 words led to the creation of 256 interaction terms. The use of multiple interaction terms increases the likelihood that our measures are reliable.

Handling negations In this usage of CASS analysis, we were concerned that the program would overlook negations. ("He is *not* liberal and he *wasn't* good tonight when he spoke.") Overlooking negations could lead to erroneous inferences. Thus, we included the facility in the software to translate common negations (not, isn't, wasn't, and aren't) into unique tokens. For example, "not liberal" would be reduced to the novel token "notliberal", without a space, thereafter treated as its own word. These adjustments increased the face-validity of the technique while leading to small changes in output; all changes in CASS effects based on this modification were less than 0.001.

Descriptive statistics of the text files To describe the text for each channel, we calculated the words per sentence, letters per word, the normative frequency of word-use, and the usage rates for the 16 target words (prior to collapsing negations). FoxNews used the fewest words per sentence ($M = 13.10$, $SD = .04$), followed by CNN ($M = 13.11$, $SD = .02$), and MSNBC ($M = 13.50$, $SD = .03$). A one-way ANOVA revealed a difference overall, $F(2, 27) = 487.36$, $p<.001$. A Bonferroni post-hoc test revealed a difference between CNN and MSNBC, $p < .001$, and between MSNBC and FoxNews, $p < .001$, but not between CNN and FoxNews. Secondly, MSNBC used the shortest words ($M = 6.29$ letters, $SD = .003$), followed by FoxNews ($M = 6.30$, $SD = .004$), and CNN ($M = 6.35$, $SD = .001$). A one-way ANOVA indicated a difference overall, $F(2, 27) = 1092.36$, $p < .001$, and post-hoc tests revealed a difference for each of the three pairwise comparisons, all ps $< .001$.

The normative frequency of words in English was measured using the HAL database, obtained from the English Lexicon Project (ELP) website (Balota et al., 2007). Due to computer memory constraints, 30 smaller text files (10 per channel for the 3 channels) were created solely for this particular analytic test. The files contained 1,000 lines randomly selected from the master corpus for each channel, producing approximately 13,000 words per file and 130,000 words per channel overall. A one-way ANOVA indicated a difference in normative frequency, $F(2, 27) = 7.10$, $p = .003$. CNN tended to broadcast less common words ($M = 9.83$, $SD = .05$), followed by MSNBC ($M = 9.87$, $SD = .02$) and FoxNews ($M = 9.88$, $SD = .02$). Post-hoc tests revealed a difference between CNN and MSNBC, $p = .03$, and between CNN and FoxNews, $p = .004$, but not between MSNBC and FoxNews. Thus, it appears that, compared to other channels, CNN used longer and lower frequency words, whereas MSNBC used longer sentences.

In Table 1, we display usage rates for the target words, channel by channel. Positive words were used about three times as often as negative words, and liberal words were used slightly more often than conservative words. Across word categories, MSNBC and FoxNews are quite comparable in their word usage rates (all differences were less than 2.0 per 10,000), whereas CNN tended to use fewer target words in every category.[1]

Statistical analyses We used non-parametric resampling analyses to statistically test hypotheses. For single-document tests, which are conducted to determine whether a single corpus shows a statistically significant mean level of media bias away from a neutral zero point as defined by the method, we used bootstrapping analyses. For a given document, 1,000 documents of equal length were generated by randomly re-sampling sentences (with replacement) from within the original text. After gleaning bootstrapped estimates of the level of bias (preferential associations) in each of these 1,000 documents, statistical significance was determined based on whether the middle 950 estimates of bias (i.e., the 95% confidence interval) included zero.

For pairwise comparisons between texts, which are conducted to determine whether media bias in one text differed from that of a second text, we employed permutation analyses. One-thousand document pairs were randomly generated by pooling all sentences across both documents and randomly assigning them to file A or B. For each of the 1,000 document pairs, we then calculated the difference in bias between documents A and B, and calculated the p value of the observed difference—for the original pair of documents—by determining its position within the distribution of permuted scores. For example, the 25th largest difference (out of 1,000) would correspond to $p = .05$, two-tailed , or $p = .025$, one-tailed.

---

[1] Our speculation is that CNN probably had lower percentages for political concepts because all of their transcripts—even apolitical ones—were more likely to be categorized by the keyword "politics" on LexisNexis (thus, the higher number of total words collected for CNN). Therefore, the CNN content was (on average) less political; this scenario would have impacted the proportions of political words listed in Table 1. CNN's neutral language (fewer valence words) could reflect greater impartiality, less value-laden reporting, or simply a lower proportion of political topics (and thus a lower proportion of value-laden programming). Language neutrality might also reflect a different programming strategy. Future research will have to sort through these speculations.

Table 1 Utterances of each target word per 10,000 words

	MSNBC	CNN	FoxNews	Mean
Positive words				
Good	12.23	11.21	12.85	12.10
Great	7.71	5.99	6.72	6.81
Best	4.24	4.71	3.59	4.18
Strong	2.56	2.39	2.43	2.46
Sum	**26.74**	**24.30**	**25.59**	**25.54**
Negative words				
Bad	3.72	3.12	3.49	3.44
Wrong	3.43	2.18	3.10	2.90
Negative	1.64	0.96	1.16	1.25
Worst	1.55	0.99	0.88	1.14
Sum	**10.34**	**7.24**	**8.63**	**8.74**
Liberal words				
Democratic	10.58	8.02	8.92	9.17
Democrats	8.56	6.37	9.36	8.10
Left	3.12	3.07	4.59	3.59
Democrat	2.27	1.34	2.08	1.90
Sum	**24.54**	**18.80**	**24.94**	**22.76**
Conservative words				
Republican	11.04	7.22	8.23	8.83
Republicans	7.10	4.58	6.72	6.14
Conservative	2.29	1.72	2.83	2.28
Conservatives	1.22	1.07	1.89	1.39
Sum	**21.65**	**14.60**	**19.67**	**18.64**

Fig. 1 Media bias as indicated by the observed 256 interactions, by channel. *Shades of blue* indicate bias in the liberal direction, and *shades of red* indicate bias in the conservative direction. These figures were created in the statistics package R with the RColorBrewer package

Results

CASS-based bias reliability A basic test of the quality of a measure is its reliability. Accordingly, we first sought to demonstrate internal consistency reliability. The channels were treated as participants and the CASS interaction values were treated as items in this analysis. The 256-item measure had excellent reliability, $\alpha = .99$.

CASS-based bias in each channel An initial analysis using original transcripts for each channel explored the observed levels of bias based on the 256 interactions. Results are displayed in Fig. 1. The distribution of correlation values is skewed to the left (blue bars) for MSNBC, whereas it is skewed to the right (red bars) for FoxNews.

Subsequently, bootstrapping analyses tested for statistically significant differences between (a) the mean level of bias (collapsing the 256 interactions by channel) and (b) the neutral zero point, where the zero point in CASS simply indicates zero preferential concept associations. MSNBC ($M = -.028$), had a liberal bias away from zero, $p < .001$. CNN ($M = -.003$) did not have a detectable level of bias away from zero, $p = .61$, and FoxNews ($M = .021$) had a conservative bias away from zero, $p = .004$. These results should be interpreted
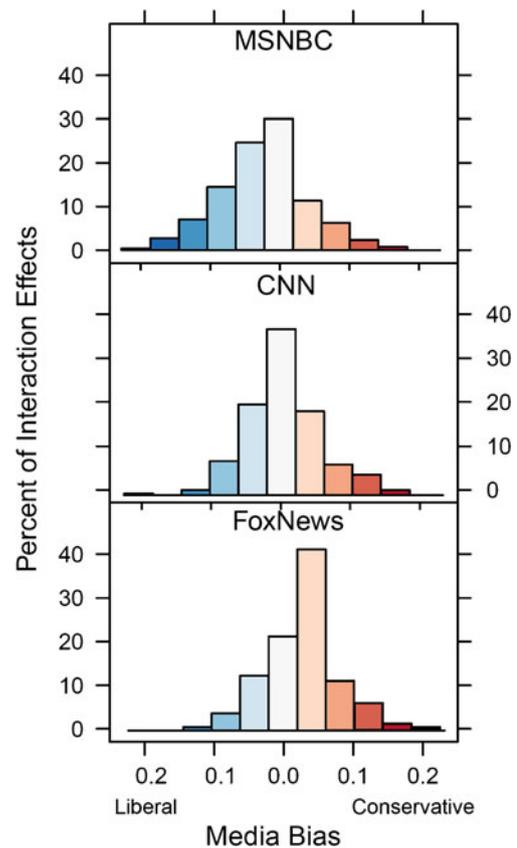
while bearing in mind the precautions outlined in the Introduction about interpreting the zero point.

CASS-based bias: channel–channel comparisons Next, we computed between-channel differences in bias, averaging over the 256 interaction terms by channel. Consistent with our prediction, the average difference between CNN and MSNBC was 0.025, reflecting a greater pro-liberal bias for MSNBC, $p = .02$. The average difference between FoxNews and MSNBC was 0.05, indicating that MSNBC exhibited significantly more liberal bias, $p < .001$. The average difference between FoxNews and CNN was 0.025, $p = .02$, indicating that FoxNews was more conservative than CNN.

CASS-based individual differences in bias: Hannity vs. Colmes Can CASS methods detect differences between specific individuals? The answer appears to be affirmative: an analysis directly contrasting Hannity's and Colmes' speech transcripts found a reliable difference between the two hosts, $M = .023$, $p = .01$; Colmes was significantly more liberal than Hannity. As revealed through the bootstrapping analyses, Hannity did not preferentially associate positive or negative concepts with liberal or conservative concepts,

$M = -.007$, $p = .32$. In contrast, Colmes did, $M = -.030$, $p < .001$. The best way to interpret these collective findings is that Colmes is relatively more liberal.

Comparing CASS to other methods: a think tank analysis Most of our results converge with popular intuitions of media biases in cable news channels (e.g., FoxNews is more conservative than MSNBC). From a practical standpoint, however, it is important to compare our measure to currently available measures. One popular measure of media bias is provided by Groseclose and Milyo (2005), who count the frequencies with which think tanks are cited. Groseclose and Milyo advocate for using both automated text-analysis and human coding. Here, we use a fully automated version of their approach. We use custom software to count the number of times that the 44 most-cited think tanks (Heritage Foundation, NAACP, etc.) were mentioned in the television transcripts. The citations were multiplied by the estimated Americans for Democratic Action (ADA) score of ideological bias (see Groseclose & Milyo, 2005, Table 1, p. 1201) of each think tank. The ADA score is an estimate of ideological (conservative/liberal) position. The output values were added and the total was divided by the number of total citations. As a simple example, imagine a particular transcript included only three citations, one citation of a think tank that has a 60.0 ADA score (output = 60) and two citations of a think tank that has a 30.0 ADA score (output = 60 total). This transcript would have a 40.0 estimated ADA score ([60 + 30 + 30]/3). As applied here, the quantitative approach is a simplified approximation of that used by Groseclose and Milyo.

Think tank analysis: reliability Distributions of scores on the outcome variables were needed to compute reliability. Therefore, sentences in the master file for each channel were randomized and then the master file was split into equal parts of ten (30 total). ADA scores were calculated for each of the ten files, for each of the three channels. The resulting internal consistency reliability was .83, indicating that the consistency of the measure was good, but somewhat lower than the reliability based on CASS.[2]

Think tank analysis: media bias in each channel The measure offered by Groseclose and Milyo (2005) involves two different ideological center-points (50.1 and 54.0 ADA scores)—different ideological levels in the United States

Congress before and after the year 1994. Importantly, the choice of a zero-point affects the inferences one draws. When the 50.1 ADA midpoint was chosen, CNN ($M = 54.28$, $SD = 1.33$) had a significant liberal bias, $t(9) = 9.96$, $p < .001$, but when the 54.0 ADA midpoint was chosen, CNN did not exhibit bias, $t(9) = 0.66$, $p = $ n.s. This ambiguity applies to think tank analyses generally: the value chosen as the null can potentially play a large role in determining one's conclusions.

Think tank analysis: channel-channel comparisons There was a main effect of channel, $F(2, 27) = 6.43$, $p = .005$. Significant differences were evident between FoxNews ($M = 50.30$, $SD = 3.19$) and CNN ($M = 54.28$, $SD = 1.33$), and between FoxNews and MSNBC ($M = 54.74$, $SD = 3.98$), both ps < .05, but there was no difference between MSNBC and CNN. Overall, the think tank analyses are largely consistent with analyses derived from CASS, yet the CASS technique was somewhat more compelling because (i) it involves more reliable measures and because (ii) there is only one zero-point indicative of zero bias.

Discussion

The present study validated a novel approach to the analysis of semantic differences between texts that builds on existing semantic space models and is implemented in novel, freely available software: CASS tools. Like all semantic space methods, CASS creates a model of the relationships among words in corpora that is based on statistical co-occurrences. The novelty of CASS is that it contrasts the semantic associations that are derived from semantic spaces, and it allows the user to compare those contrasts across different semantic spaces (from groups or individuals).

The application of CASS tools to the domain of media bias replicated, or is consistent with, previous findings in this domain (Groeling, 2008; Groseclose & Milyo, 2005). Moreover, direct comparison with an existing method for detecting political bias produced comparable but somewhat stronger results while virtually eliminating the need for researchers to make choices (e.g., about which Congress provides the zero) that could ultimately influence the researchers' conclusions.

More generally, semantic space models such as BEAGLE provide several key advances toward the objective study of semantic structure. First, the model captures indirect associations among concepts in the media—effects that may influence viewers' ideology (Balota & Lorch, 1986; Jones & Mewhort, 2007; Landauer & Dumais, 1997). For example, using the name of Jeremiah Wright (Obama's pastor), an association chain such as democrats-Wright-bad would lead to a slight bias against democrats because "democrats" becomes slightly associated with "bad". These indirect associations may

---

[2] This analysis essentially used only ten items per channel, in comparison to the 256 items used in the CASS reliability analysis. Thus, we wanted to make sure that CASS still produced relatively higher reliability when it too was based on ten split files, based on the collapsed output of the 256 interaction scores. Indeed, this ten-item measure when applied to CASS still produced excellent reliability, $\alpha = .94$, indicating that the CASS approach has superior reliability when under similar constraints.

influence cognition outside of awareness (i.e., implicit media bias), thus limiting people's ability to monitor and defend against propaganda effects, particularly as the association chains become longer and more subtle. Comprehensive analysis of indirect effects would be tedious if not impossible to conduct manually. Although clearly human coders would be better at detecting some characteristics of lexical information (e.g., non-adjacent negations; "not a bad liberal"), for information like indirect associations it is superior to use unsupervised machine learning via CASS and its parent program BEAGLE. The nuance that CASS detects can provide a complement to other measures of subtle lexical biases (Gentzkow & Shapiro, 2006), and it can increase the likelihood of capturing bias.

Limitations One limitation of our approach to the measurement of media bias is its dependence on lexical information. Media bias is likely to be communicated and revealed in multiple ways that cannot be detected solely by using lexical analysis (e.g., sarcasm, facial expressions, and pictures on the screen). We posit that the effect sizes we found are probably *underestimated* due to our sole reliance on lexical information. It would be interesting to combine or compare our lexical approach with a non-lexical one (Mullen et al., 1986). Second, the finding that CNN showed no reliable bias away from zero should not necessarily be taken to imply that CNN reports on politics in an objective manner; rather, the implication is simply that CNN does not associate liberal or conservative words preferentially with positive or negative words. This type of limitation also applies to other metrics of political bias, many of which must (a priori) specify the midpoint—whereas, fortunately, CASS does not.

Summary and conclusions Advances in the use of semantic space models have yet to reach their full potential. With the goal of enhancing the utility of such models, tools for Contrast Analysis of Semantic Similarity (CASS) provide the novelty of contrasting associations among words in semantic spaces—resulting in indices of individual and group differences in semantic structure. Thereafter, the tools can compare the contrasts obtained from different documents. The tools can be used to study many types of sample units (e.g., individuals, political bodies, media outlets, or institutions) and many topics (biases, attitudes, and concept associations). Specific topics may include racism (black, white, good, bad) and self-esteem (I, you, good, bad); topics that do not involve valence could be explored with this approach as well, such as stereotyping (black, white, athletic, smart), or self-concept (me, they, masculine, feminine). Ultimately, we hope readers view our software as a practical package for creating new semantic spaces that help them extract individual and group differences from their own streams of text.

References

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*(3), 533–548.

Balota, D. A., & Lorch, R. F. J. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 12*, 336–345.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The english lexicon project. *Behavior Research Methods, 39*, 445–459.

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality, 42*(1), 96–132.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. [Proceedings Paper]. *Personality and Social Psychology Bulletin, 30*(10), 1332–1346.

Gentzkow, M. A., & Shapiro, J. M. (2006). What drives media slant: Evidence from daily newspapers. NBER Working Paper Series

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Groeling, T. (2008). Who's the fairest of them all? An emprical test for partisan bias on ABC, CBS, NBC and Fox News. *Presidential Studies Quarterly, 38*, 631–657.

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics, CXX*, 1191–1237.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*, 1–37.

Klebanov, B. B., Diermeier, D., & Beigman, E. (2008). Lexical cohesion analysis of political speech. *Political Analysis, 16*, 447–463.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Lund, K., & Burgess, C. (1996). Producing high-dimensional spaces from lexical co-occurence. *Behavior Research Methods, Instruments, & Computers, 28*, 203–208.

Mullen, B., Futrell, D., Stairs, D., Tice, D. M., Baumeister, R. F., Dawson, K. E., et al. (1986). Newscasters' facial expressions and voting behavior of viewers: Can a smile elect a president? *Journal of Personality and Social Psychology, 51*, 291–295.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2006). Website: https://implicit.harvard.edu/implicit/.

Pennebaker, J. W., & Chung, C. K. (2008). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorff & M. A. Bock (Eds.), *A content analysis reader* (pp. 453–465). Thousand Oaks, CA: Sage.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Erlbaum.