Measures of Argument Strength

A Computational, Large-Scale analysis of Effective Persuasion in Real-World Debates

Youk, S.; Malik, M.; Chen, Y.; Hopp, F.R.; Weber, R.

DOI
10.1080/19312458.2023.2230866

Publication date
2024

Document Version
Final published version

Published in
Communication Methods and Measures

COMMUNICATION METHODS AND MEASURES
2024, VOL. 18, NO. 1, 7–29
https://doi.org/10.1080/19312458.2023.2230866

Routledge
Taylor & Francis Group

Measures of Argument Strength: A Computational, Large-Scale Analysis of Effective Persuasion in Real-World Debates

Sungbin Youk[a], Musa Malik[a], Yibei Chen[a], Frederic R. Hopp[b], and René Weber[a,c,d]

[a]Department of Communication - Media Neuroscience Lab, University of California, Santa Barbara, USA; [b]School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands; [c]Department of Psychological and Brain Sciences, University of California, Santa Barbara, USA; [d]School of Communication and Media, Ewha Womans University, Seoul, South Korea

ABSTRACT

The present research examined how value-free and value-driven measures of argument strength (MAS) can be computationally extracted using a theory-driven approach at scale in a naturalistic setting by analyzing a total of 7,961 real-world debates and 42,716 judgments in rhetorical quality. In the first study, value-free MAS was significantly related to the rhetorical quality of arguments (i.e. their persuasiveness). The results indicate that the side that provides more information-source citation, less quantitative specificity, more unique words, and more abstract language is more likely to be perceived as convincing in dialectical argumentation, where two people are exchanging opposing arguments. In the second study, the added influence of value-driven MAS is investigated. The results show that the similarity between the moral values represented in arguments and those that are salient to argument receivers predicts the rhetorical quality. The research demonstrates how rhetorical quality can be measured and predicted at scale, and how naturally generated arguments can be used for scientific progress in persuasion research.

Arguments are a universal aspect of human social interaction and have persisted throughout the history of mankind: from political and philosophical debates in ancient Greece to contemporary online debate forums, such as *Change My View*,[1] *Debate.org*,[2] and *Kialo.com*.[3] Even without formal training, albeit with varying degrees, individuals can use their innate, critical thinking when evaluating an argument (H. Hoeken et al., 2012), highlighting humans' functionally evolved capacity to use argumentation for communication (Mercier & Sperber, 2011). The literature on effective (i.e., successful) and ineffective argumentation and persuasion is vast, spanning from popular literature (e.g., Cialdini, 2008a; Sinnott-Armstrong, 2018), to scholarly textbooks (e.g., Dillard & Pfau, 2002; O'Keefe, 2002), to countless articles published in scientific journals (e.g., Carpenter, 2015; H. S. Park et al., 2007). While this literature elucidates many argument characteristics that individuals prioritize when judging the strength of an argument, to date there is no reliable, valid, and accessible procedure available for extracting indicators of argument strength from text that meet the following three criteria: independence from self-reports, scalability, and ecological validity. By measuring argument strength independent from self-reports, we can avoid the tautological nature of identifying an argument as effective when another group of individuals labeled an argument as strong. As self-

CONTACT René Weber renew@comm.ucsb.edu Department of Communication - Media Neuroscience Lab, University of California, Santa Barbara, CA 93106-4020, USA
[1]https://www.reddit.com/r/changemyview
[2]https://www.debate.org
[3]https://www.kialo.com
Supplemental data for this article can be accessed online at https://doi.org/10.1080/19312458.2023.2230866

reports include subjectivity and idiosyncratic biases, it is erroneous to assume that argument evaluation by a group of individuals will be equivalent to another group's evaluation unless we introduce a large amount of control variables. Additionally, some studies failed to differentiate argument strength and argument effect in self-reports: self-reported argument strength evaluations were included in the measurement of perceived argument effectiveness (e.g., H. Hoeken, 2001). As for scalability and ecological validity, we have yet to develop a systematic and theoretically driven procedure of measuring argument strength that enables researchers to study argument strength in large amounts of text produced by individuals arguing with each other in naturalistic settings, and not in artificially created, manipulated experiments. We call indicators that meet these criteria Measures of Argument Strength (MAS).

Developing a better understanding of the characteristics that drive argument strength unrelated to self-reports, in naturalistic settings, and at scale is instrumental for several reasons. First, MAS are inherently independent of subjectivity and scalable via a computational, non-human evaluation process. As such, MAS are better suited for generalizing to large and diverse groups of communicators, which should increase their potential to predict instances of successful argumentation in naturalistic, large-scale (online) discussions. Second, studying argumentation and persuasion in large, diverse groups should be beneficial for advancing persuasion theories and research on attitude-change, which so far have largely been studied in controlled laboratory settings using WEIRD (Western, Educated, Industrialized, Rich, Democratic) samples, such as undergraduate students in the United States (Rad et al., 2018; Zhao & Cappella, 2016; Zhao et al., 2011). Third, there has been a recent increase in the popularity of online discussion forums, which have the potential to shape public opinion and even influence trends in online polarization (e.g., Wang et al., 2018).

The presented research conducts two studies to develop and validate procedures to automatically extract two types of MAS from text (i.e., value-free and value-driven MAS), and to make these procedures accessible to communication scholars. The two studies use a large text corpus from the online debate platform *Debate.org* that includes a total of 7,961 real-world debates and 42,716 judgments on argument strength.

Persuasive arguments

Argumentation theory is an umbrella term that encompasses the study of production and evaluation of argumentation. Argumentation is similar to formal logic (e.g., deductive logic and implication) as they both consider an argument to include reasoning of some premise and a logically derived conclusion (Hitchcock, 2006). The dialectic nature (i.e., the exchange of contrasting viewpoints) of argumentation is what separates them (Blair & Johnson, 1987; van Eemeren et al., 2009). Thus, the evaluation of an argument is relative and contingent upon the argument itself and the opposing argument (Habernal & Gurevych, 2016). This study focuses on examining the rhetorical quality of arguments (see Wachsmuth, Naderi, Habernal, et al., 2017 for other types of argument quality). Evaluating rhetorical quality assumes that the purpose of an argument is to persuade the audience to support or adhere to a specific stance of a given debate (Blair, 2012; Mercier & Sperber, 2011). This is consistent with how Mercier and Sperber (2011) conceptualize reasoning: the "mental action of working out a convincing argument, the public action of verbally producing this argument so that others will be convinced by it, and the mental action of evaluating and accepting the conclusion of an argument produced by others" (p. 59). In the following sections, we are using the terms rhetorical quality, argument quality, and argument strength interchangeably.

There are various sources or predictors of rhetorical quality, such as ethos, pathos, and logos, according to the Aristotelian perspective (McCormack, 2014), and the communicator, message, medium, and recipient in Lasswell's model of communication (Lasswell, 1948). Considering how these predictors may have an interactive effect (e.g., the interaction between audience characteristics and argument strength in the dual-processing model; Petty & Cacioppo, 1986), it is important to decompose rhetorical qualities that are derived from intrinsic characteristics of the argument and from

the individual differences of the audiences. Therefore, study 1 focuses on computationally extracting argument characteristics to predict successful persuasion that are largely independent of individuals' value systems (i.e. value-free MAS). In study 2, we investigate the added and interactive influence of receivers' moral value systems in relation to the moral content of the argument (i.e., value-driven MAS).

Study 1

Value-free MAS

An extensive body of work in persuasion research has identified numerous characteristics of arguments or messages that can predict argument strength (e.g., whether an argument is construed as one-sided or considers multiple perspectives; for an overview, see Cialdini, 2008b; O'Keefe, 2002). However, many of these characteristics require separate extrinsic evaluations by humans, including self-reports of argument receivers (e.g., an audience or target group of an argument) and content analysis by human annotators (Baesler & Burgoon, 1994). Besides relying on the judgment of a small group of people, examination of the relationship between argument characteristics and quality has also commonly involved researchers creating unsystematic message variations for experiments (O'Keefe & Jackson, 1995). These procedures persist in persuasion research (e.g., C. Y. Li, 2013; Stavraki et al., 2021; Wall & Warkentin, 2019; Yi et al., 2013). In comparison, the argumentation mining literature has also examined the relationship between argument characteristics (i.e., linguistic characters) and quality (e.g., H. Li et al., 2020; Luu et al., 2019; Tan et al., 2016). These studies included a large sample of argument quality judgments in a more naturalistic setting, such as an online debate. However, they fall short in their theoretical conceptualization, especially in explicating the rationale that explains why certain argument characteristics affect rhetorical quality.

To bridge the gap in the persuasion and argumentation mining literature, we propose a non-exhaustive list of five argument characteristics that predict argument quality based on these six criteria: (1) the relationship has theoretical relevance in persuasion literature; (2) the characteristics can produce substantial variance across different arguments; (3) the variations in characteristics can be evaluated by an automated text analysis (i.e., does not involve human annotators); (4) the automated text analysis tools are validated and easily accessible to communication scholars; (5) the characteristics can predict persuasiveness of the argument that is evaluated by a large, diverse group of individuals; (6) the predictions are independent of argument receivers' individual value systems.

Information-source citation

Information-source citation refers to overtly stating the origin of the evidence that supports an argument. According to O'Keefe (1998), using information-source citation is one of the ways of making an argument more explicit. An explicit argument increases the rhetorical quality of an argument as providing supporting evidence signals that the one who produced the argument is honest and well-informed. O'Keefe's meta-analysis of 23 studies indicated that citing the source of the information was positively related to the persuasiveness of the argument, $r = .06$, 95% CI = (.01, .11).

Hyperlinks, a blue underlined text which embeds a reference to be accessed with a click of a button, is an easy way of citing the source in text-based arguments, such as an online debate platform, and its persuasive effect has been broadly researched (H. Park & Thelwall, 2003). Studies on *ChangeMyView* (a subreddit where people share opposing viewpoints) have found that the number of hyperlinks are positively correlated with the number of replies to the argument (Tan et al., 2016) and attitude change (Priniski & Horne, 2018). According to news framing and credibility research (e.g., Borah, 2014; K. A. Johnson & Wiedenbeck, 2009), hyperlinks enhance the perceived credibility of the information and news story. Increasing the credibility appeals to ethos, which consequently increases argument quality (McCormack, 2014). Therefore, this study predicts that the use of hyperlinks in arguments

boosts argument quality. It should be noted that the quality of evidence in the cited source is not examined in this study (see limitations for further discussion).

H1: The more information-source citations an argument has, the higher its rhetorical quality.

Quantitative specificity

Quantitative specificity refers to providing specific quantitative values (e.g., percentages) of evidence. For example, "*delays on services run by Transport for London have increased by two-folds since 2013 because of overcrowding*" has quantitative specificity as it explicitly provides numeric information: two and 2013. In contrast, "*delays on service run by Transport for London have increased substantially from the past*" is not quantitatively specific. Providing quantitative information indicates not only that there is evidence to support the argument, but that the support is derived from measurable data. This may imply that reliable evidence corroborates the argument. Using statistical evidence is an example of quantitative specificity. For instance, according to H. Hoeken and Hustinx (2009), anecdotal evidence (which lacks quantitative specificity) is less persuasive than providing statistical evidence.

In contrast to the rationale supporting the relationship between quantitative specificity and argument quality, exemplification theory states that an individual's perceived argument quality will be more affected by examples, not numbers (Zillmann, 1999). Examples grab attention, facilitate cognitive processing, and trigger emotional experience, which can enhance an argument's persuasive impact. According to Zillmann (1999) and Brosius (2000), as lay people do not utilize quantitative information in everyday situations, they may be unfamiliar with comprehending, incorporating, and utilizing such information when evaluating an argument.

The empirical evidence surrounding the role of quantitative specificity in argument persuasion effects is inconclusive. A meta-analysis done by O'Keefe (1998) found a statistically non-significant correlation between the use of quantitative specificity and argument quality. However, O'Keefe's meta-analysis included only 8 studies, which suggests that this MAS has been largely understudied, or put differently, that the meta-analytical finding is not based on strong evidence from sufficiently replicated original studies. In addition, the studies included in O'Keefe's meta-analysis examined quantitative specificity by comparing a limited number of arguments (usually two) that are purposefully created by the researchers. The effect of quantitative specificity on argument quality may be more pronounced when examined in the naturalistic environment of an online debate platform and at scale. To further examine the impact of quantitative specificity, the following research question is asked:

RQ1: How is the use of quantitative specificity related to the rhetorical quality of an argument?

Argument length

Argument length, which is also referred to as argument quantity, may be positively related to its rhetorical quality. More specifically, persuasive arguments tend to have more words, sentences, and paragraphs compared to those that are less persuasive (Tan et al., 2016). Among the various linguistic features that were examined in Durmus and Cardie's (2019) study of online debates, the length of an argument consistently improved the prediction of argument quality across various debates. This may be because argument length potentially reflects the amount of information within the argument that has the potential to make it convincing.

It should be noted, however, that overly lengthy arguments may also reduce their persuasiveness because lengthy, unclear, and complex arguments require more cognitive and motivational resources (Kruglanski & Thompson, 1999; Pierro et al., 2005). Additionally, as emphasized by Hornikx and Hoeken (2007), merely adding weak and unstructured premises to increase the length of an argument does not increase argument quality. Yet, as shown by Heit and Rotello

(2012), when arguments were logically invalid, they were considered to be more persuasive when they were longer. For logically valid arguments, longer arguments were considered to be less persuasive. Consequently, the relationship between argument length and rhetorical quality may be complex, or at least nonlinear; the available evidence does not warrant a directional hypothesis, and thus we ask:

RQ2: How is the length of an argument related to its rhetorical quality?

Argument sentiment

It is natural for people to mix rational and emotional judgment in argumentation (Villata et al., 2017). Emotions in an argument carry information value and are considered when evaluating the quality of the argument (Gilbert, 2004). To operationalize emotions in an argument, this study focuses on argument sentiment: the degree to which an argument adopts positive (e.g., good, great, and nice) or negative (e.g., bad, awful, and horrible) language. This aligns with how previous studies have also conceptualized emotion and sentiment in arguments (e.g., H. Li et al., 2020).

By conceptualizing argument sentiment as the emotional valence expressed by the content of an argument, this study draws upon research on attitude change as a result of persuasive arguments. The majority of research regarding attitude change has assessed the effects of argument sentiment on receivers' emotions and demonstrated the existence of the well-known negativity bias (Cacioppo et al., 1997). The negativity bias suggests that individuals tend to focus longer on negative information and weigh negative information more heavily during decision-making (Fiske, 1980). Additionally, arguments that prompt negative emotions are likely to be remembered and have a lasting impact on attitude change (Nabi, 1999). The negativity bias was observed in Tan et al. (2016) that examined the persuasiveness of arguments on the online debate platform *ChangeMyView*. They found that persuasive arguments tend to have negative sentiments. Given the available theorizing and evidence, the following hypothesis is proposed:

H2: The more negative the sentiment of an argument, the higher is its rhetorical quality.

Argument concreteness

Argument concreteness considers the degree to which viewpoints are fully and specifically articulated: "Evasion, concealment, and artful dodging (. . .) are and should be excluded from an ideal model of critical discussion" (van Eemeren et al., 1993, p. 173). O'Keefe (1998) states that concrete arguments are more persuasive than arguments that are abstract, as their logic and support are clearly conveyed. At the same time, it is possible that abstract arguments are superior regarding rhetorical quality, as they provide fewer opportunities to be countered by opposing evidence or by revealing specific flaws in their logic (Sinnott-Armstrong, 2018).

Although argument concreteness is conceptualized at the argument level, it can also be derived from a word-level analysis. According to logical atomism, the complex meaning of language can be broken down into its most elementary units of meaning: words (Kim, 1972). An argument that is less concrete is more likely to use words that are abstract and do not refer to a perceptible entity. A concrete argument is likely to entail direct references to perceptible entities. In the current study, argument concreteness is conceptualized as the degree to which an argument contains words that are either abstract in that they do not refer to a perceptible entity (e.g., democracy, government, law, etc.), or that are concrete and thereby refer to perceptible entities (e.g., banana, raven, machine gun, etc.). To examine the relationship between rhetorical quality and argument concreteness, we ask the following research question:

RQ3: What is the association between the concreteness of words in an argument and its rhetorical quality?

Methods

Using a purpose-built Python script, 24,668 debates rated by 4,582 unique users were collected in October, 2021 from *Debate.org*, which is an online debate platform that is no longer available since June 2022[4] due to low website traffic. The collected debates were posted between January 2009 and September 2017, and span across 23 different debate categories as identified by the platform (e.g., politics, religion, philosophy, science, education, and entertainment). Although some users chose not to disclose their personal and demographic information, there are substantial variances in the users' gender, age, income, education, ideology, political affiliation, and religion (see Supplement Materials for more information on *Debate.org* and descriptive statistics on users' profiles).

For each debate, two users exchange their opposing perspectives as the instigator or the contender: instigator and contender are the argument producers. The instigator and the contender take opposing stances (i.e., pro- vs. contra-stance) to the given debate. Each exchange (i.e., a set of pro-stance argument and a set of contra-stance arguments) constitutes a single round. The debate can extend to multiple rounds: for the data analyzed in this study, a debate had 3.57 rounds on average (Q1 = 3, Q2 = 3, Q3 = 4, $SD = 48.01$). Other users serve as argument receivers and evaluate the two argument producers based on several criteria, including the convincingness of the arguments. We collected information about the debates (i.e., the instigator and contender's user ids, their stance, and arguments), the evaluations or votes (i.e., the audience user's id and the voting outcomes), and the audiences' debate history (i.e., all arguments made by the users or argument receivers who voted on a debate).

Rhetorical quality

On *Debate.org*, argument audiences evaluate the debate on seven criteria: who they agreed with before the debate, who they agreed with after the debate, who had better conduct, who had better spelling and grammar, who used the most reliable sources, and who made more convincing arguments. It should be noted that users voted on the level of the entire debate (not on each round or each argument). Therefore, we analyze rhetorical quality (and the following MAS) on a debate-level, which encompasses all the arguments made by the two opposing sides.

In line with the conceptualizations discussed above, we operationalized rhetorical quality by focusing on one of the seven criteria: a forced-choice vote on which side (pro or contra) made the more convincing argument. This operationalization is consistent with Habernal and Gurevych's (2016) that arguments are evaluated against their opposing argument. Accordingly, rhetorical quality was dummy coded for each vote (0 = contra-stance made the more convincing argument, and 1 = pro-stance made the more convincing argument).

Information-source citation

For each of the arguments made in a debate, the number of citations was measured by counting the number of hyperlinks in an argument. The hyperlinks are embedded in the argument text as either "https://"or"." The string-matching function in the Python Regex library[5] was used to count the number of hyperlinks in an argument. The number of citations used in the contra-stance arguments

[4]web.archive.org has archived some pages of the website. Here is an example of a debate: https://web.archive.org/web/20220319231240/https://www.debate.org/debates/Intelligence-Nature-vs-Nurture/1/

[5]https://docs.python.org/3/library/re.html

was subtracted from the number of citations used in the pro-stance arguments. Thus, a positive score indicates that the arguments from the pro-stance contained more hyperlink citations compared to the arguments from the contra-stance. The information-source citation measure ranged from −161 to 87 (Q1 = −2, Q2 = 0, Q3 = 1) with a mean of −0.01 (*SD* = 5.86).

To validate the relevance of the hyperlinks (i.e., the hyperlinks serve the purpose of providing information to support the argument), we conducted a manual annotation with the help of 10 trained human coders. The human coders rated the arguments on a 4-point scale (1 = none of the hyperlinks are related to making the argument, and 4 = all the hyperlinks were related to making the argument). After achieving acceptable agreement (Cohen's Kappa above .8), each coder independently annotated a set of arguments. In total, a random sample of 1,247 (7.8% of the arguments in study 1 were validated. The average score was 3.22, indicating that most of the hyperlinks were relevant to the argument. Only 11.9% of the arguments had irrelevant hyperlinks.

Quantitative specificity

The use of numbers or numeric words in an argument was counted computationally. The text in each argument was preprocessed by eliminating numbers that were used for structuring the argument (e.g., *Round 1*, *[2]*, and *9:*) and replacing numerical words with numerical digits (e.g., three thousand into 3000) using word2number library.[6] After tokenizing the text of each argument, the number of numerical tokens in the arguments made by the contra-stance was subtracted from that of the arguments made by the pro-stance. Again, a positive score indicates that the pro-stance arguments within the debate contained more quantitative specificity compared to the contra-stance argument. The quantitative specificity measure ranged from −525 to 8,414 (Q1 = −4, Q2 = 0, Q3 = 5, *M* = 1.09, *SD* = 48.01).

Length

After removing punctuation and stopwords, lowercasing, and stemming (which refers to identifying the root form of the word), a list of unigram tokens (i.e., tokens that are considered being independent of the tokens before it) was obtained for each argument. We constructed a list of unique unigram tokens for each stance in a debate. The number of unique unigram tokens is used to control for use of repetition since we conceptualized the effect of argument length in terms of elaboration, complexity, and structure of the argument. We operationalized argument length by subtracting the number of unique tokens in the contra-stance argument from the pro-stance argument. A positive score indicates that the pro-stance arguments contained more unique tokens compared to contra-stance arguments. The measurement of length ranged from −1,016 to 2,714 (Q1 = −82, Q2 = −5, Q3 = 77) with a mean of 0.39 (*SD* = 162.26).

Sentiment

The sentiment of the argument was computed by using the Natural Language Toolkit's[7] Valence Aware Dictionary for sEntiment Reasoning[8] (VADER; Hutto & Gilbert, 2014) Python library. VADER was chosen among sentiment dictionaries as it uses a rule-based model and outperforms other dictionaries (Bonta et al., 2019; Hutto & Gilbert, 2014). VADER provides a continuous rating of sentiment for over 9,000 words. We calculated the sentiment for each stance of the debate: a compound sentiment score is provided by running the polarity_scores function on the arguments that are aggregated at the stance level. The sentiment score ranges from −1 (very negative sentiment) to + 1 (very positive sentiment). We calculated the difference in the sentiment score for the two

[6]https://pypi.org/project/word2number/
[7]https://www.nltk.org/
[8]https://github.com/cjhutto/vaderSentiment

opposing stances (*min* = −2.00, *max* = 2.00, Q1 = −0.10, Q2 = 0, Q3 = 0.23, *M* = 0.06, *SD* = 0.93). A positive score indicates that pro-stance arguments that are made within the debate contained more positive sentiment compared to contra-stance. For instance, the sentiment of the contra-stance arguments was extremely negative (sentiment score of −1) while the sentiment of the pro-stance arguments was slightly negative (sentiment score of −0.2). The difference score is + 0.8: a positive score indicates that the pro-stance arguments are relatively more positive (i.e., relatively less negative) than the contra-stance arguments.

Concreteness

To compute the concreteness of each argument, Brysbaert et al. (2014) corpus, which contains concreteness ratings for 40,000 generally known English word lemmas, which are the base or dictionary form of a word, was used. Each tokenized lemma in an argument was compared to the concreteness rating in Brysbaert et al. (2014)'s dictionary (1= very abstract and 5 = very concrete). The concreteness scores were averaged across all the arguments made by each stance of the debate. The difference in the average concreteness scores between the pro-stance and the contra-stance argument was measured (*min* = −3.48, *max* = 5, Q1 = −0.05, Q2 = 0, Q3 = 0.07, *M* = 0.01, *SD* = 0.22). A positive score indicates that the pro-stance arguments contained more concrete words compared to the contra-stance arguments. The preprocessed data and the notebook for computationally extracting value-free MAS are available in our OSF repository.[9]

Results

As the votes were nested in debates, we conducted a mixed-effects binomial logistic regression, which enables us to robustly analyze rhetorical quality (first level) across debates (second level). As this study focuses on examining the effects of value-free MAS on rhetorical quality (i.e., fixed effects at the first level), the small cluster size (i.e., the number of votes within a debate) does not lead to serious bias; hence, some researchers set the minimum sample size per cluster as 1 (Bell et al., 2008; Clarke, 2008; Clarke & Wheaton, 2007; Maas & Hox, 2005). However, to assure the convergence of the model, we analyzed the debates that had more than 2 votes. In the 24,668 debates that we collected, there were a total of 71,788 votes and an average of 2.49 votes within a debate (*min* = 0, *max* = *100, Q1 = 1, Q2 = 2, Q3 = 3, SD = 2.93*). Consequently, we analyzed 42,716 votes on 7,961 debates in this study.

Before examining the relationship between the effect of value-free MAS and rhetorical quality, we first constructed an unconditional mean model (see Model 1 in Table 1) to examine debate-level variations in rhetorical quality without any additional predictors (i.e., fixed effects). This model shows that the 61% of the variance in rhetorical quality can be explained by the debate level grouping.

To examine how value-free MAS predicts rhetorical quality after accounting for the debate-level variance, the predictors were standardized and added to the model as fixed effects (see Model 2 in Table 1). Multicollinearity was not an issue as the largest variance inflation factor (VIF) was less than 1.45. The results shows that the addition of value-free MAS significantly improves the model, X^2 (5) = 2693.8, $p < .001$. The fixed effects alone explain 24% of the variance in rhetorical quality. Consistent with H1, the stance that provided more information-source citations (odds ratio = 1.47) was more convincing. The stance with a more convincing argument also was less likely to use quantitative specificity (odds ratio = 0.60), more likely to be longer (odds ratio = 3.87), and less likely to be concrete (odds ratio = 0.81), which answers RQ1, RQ2 and RQ3, respectively. As for H2, the results indicate that the relationship between argument quality and use of negative words is not statistically significant. In other words, providing one standard deviation more hyperlinks (i.e., 5.86 more links) than the opposing side increases the odds of being identified as more convincing by 47%. Using less quantitative specificities by one standard deviation (i.e., 48.01 more) increases the odds of being more persuasive by 40%. One standard deviation increase in length of the argument (i.e., 162.26 more unique words) than the opposing side increases the odds of having

[9]https://osf.io/ax5mg/?view_only=c31b414de0544a62be3d2e97aa18e4e3

Table 1. Summary of mixed-effects binomial logistic regression for value-free MAS (Study 1 and Study 2).

Fixed Effects	Study 1: Model 1	Study 1: Model 2	Study 2: Model 3	Study 2: Model 4
(Intercept)	0.57***[0.54–0.60]	0.58***[0.55–0.62]	0.52***[0.52–0.52]	0.54***[0.51–0.58]
Information-source citation		1.47***[1.39–1.55]		1.49***[1.39–1.60]
Quantitative specificity		0.60***[0.51–0.71]		0.65***[0.54–0.78]
Length		3.87***[3.59–4.17]		4.55***[4.13–5.01]
Sentiment		0.95 [0.91–1.00]		0.93* [0.87–0.99]
Concreteness		0.81***[0.77–0.86]		0.76***[0.70–0.81]
Random Effects				
σ^2	3.29	3.29	3.29	3.29
τ_{00}	5.08	3.41	6.01	4.00
ICC	0.61	0.51	0.65	0.55
Marginal R^2		0.24		0.27
Conditional R^2	0.61	0.63	0.65	0.67
Number of debates (*k*)	7,961	7,961	5,990	5,990
Number of votes (*n*)	42,716	42,716	27,534	27,534

* $p < .05$, ** $p < .01$, *** $p < .001$. For the fixed effects, the odds ratios are provided and the 95% confidence intervals using a Wald z-distribution approximation are provided in brackets. All predictors are standardized.

a better argument quality by more than three times. One standard deviation increase in the concreteness of the words in the argument (i.e., 0.22 scores higher) compared to the opposing side decreases the convincingness likelihood by 20%.

Discussion

Contrary to the idea that there is a discrepancy between how argumentation theorists conceptualize a persuasive argument and how laypeople evaluate argument quality (O'Keefe, 1995), this study has identified four salient value-free MAS (i.e., information-source citation, quantitative specificity, length, and concreteness) that are independent of self-report, theoretically driven, and computationally measured at scale that can predict people's judgment in rhetorical quality. Sentiment was not a salient predictor in this study.

In dialectic argumentation, the side that provides longer arguments, more citations for the source of the information is likely to be more convincing because these measures are related to the quality and quantity of supporting evidence. When the argument spells out the supporting details and provides additional explanation, it tends to be longer (O'Keefe's, 1998; Wachsmuth, Naderi, Habernal, et al., 2017). By providing more links to the source of the evidence, the argument gains credibility. The argument receivers and audiences can also directly access the source with ease using the hyperlinks, providing further evidence in favor of the argument.

As for quantitative specificity, the results were inconsistent with O'Keefe's (1998), which found no significant effects. At least for *Debate.org*, a naturalistic environment that provides a large number of arguments with great variance of quantitative specificity, the stance that uses less quantitative specificity is more likely to be perceived as more convincing. This finding is consistent with Zillaman's exemplification theory.

The study also found that the side using more abstract words in its arguments was more persuasive. This is inconsistent with previous literature that found a positive relationship between the use of concrete language and persuasiveness (Hansen & Wänke, 2010), comprehensibility (Sadoski et al., 2000), and the perceived importance of the argument (Miller et al., 2007). As stated above, abstract arguments are difficult to argue against (Sinnott-Armstrong, 2018). In *Debate.org*, two argument providers with opposing views exchange arguments throughout multiple rounds, giving them several opportunities to rebut each other. As argument receivers and audiences evaluate the argument quality at a debate-level by comprehensively considering the entire dialectic exchange and choosing whether the pro-stance or the contra-stance made a more convincing argument, the argument receivers may consider the argument stance to be less convincing when it did not refute the opposing arguments. At the same time, it is possible

that abstract language is associated with higher-order values, such as moral values. Arguments that appeal to moral values (i.e., fairness and loyalty) are likely to use abstract language as they are abstract concepts. However, the relationship between moral language, abstract language, and the receiver's evaluation of argument quality cannot be examined with value-free MAS. This requires a value-driven MAS.

In this study, sentiment was not a significant predictor of argument quality. The dialectic structure of the debate may have attenuated its effect on rhetorical quality. The contra-stance is likely to use more negations and refutations to argue against the provided debate topic. To empirically demonstrate the relationship between the stance and sentiment, we conducted a dependent sample *t*-test by comparing the sentiment score of the argument made by pro- and contra-stance. The post hoc analysis reveals that arguments made by the contra-stance are less likely to be positive ($M = 0.09$, $SD = 0.87$) than the pro-stance ($M = 0.15$, $SD = 0.86$), $t = 12.29$, $p < .001$).

To further anatomize argument quality, we examine the effect of value-driven MAS in the next study. As certain moral frames of an argument can affect perceived argument quality based on the argument receivers' moral values (e.g., Aramovich et al., 2012; Luttrell et al., 2017; Weber et al., 2015), the value-driven MAS focuses on moral framing of the arguments and moral value systems of the argument receivers. We will also explore the interaction between value-free and value-driven MAS because value-driven MAS may provide a systematic, computational, scalable measure for capturing the variance in value-free MAS across various debate topics and issues.

Study 2

Study 1 focused on MAS (i.e., value-free MAS) that capture primarily linguistic characteristics of arguments, which clearly is not the only information that argument receivers and audiences consider when voting on which particular argument is more convincing to them. Rather, arguments are received and evaluated against audiences' a priori held attitudes and value systems (Boote, 1981; Shen & Edwards, 2005; Watt et al., 2008). For instance, an argument produced by contenders and instigators pro or counter to the constitutional right of abortion in the US, which is identical in its linguistic characteristics, will be voted on very differently regarding rhetorical quality in liberal versus conservative audiences. As individuals of an audience hold various a priori attitudes that are integrated in many different value systems, it is difficult, if not impossible, to study the effect of value-driven MAS and their interaction with the value-free MAS in study 1. A general framework for investigating value-driven MAS is needed as well as narrowing the focus of study to a value system with general relevance for a multitude of arguments and debates. We attempt to meet these requirements by drawing on Bench-Capon's (2003) Audience-specific Value-based Argumentation Framework (AVAF) and by focusing on a moral-value system as defined by Moral Foundations Theory (MFT; Haidt, 2007).

Audience-specific value-based argumentation framework

The Audience-specific Value-based Argumentation Framework (AVAF; Bench-Capon, 2003) explains why one argument may be preferred over another based on the values advocated by the arguments and the values that are upheld by the audiences. AVAF can be formalized as a quintuple:

$$\text{AVAF} = \langle Ar, att, V, val, \succ_{\alpha} \rangle$$

Ar is a finite set of arguments; *att* is an irreflexive binary relation on an argument, which demonstrates the relationship between arguments: either opposing (i.e., attacking) or consistent (i.e., not attacking); *V* is a nonempty set of values, such as social and moral values; *val* maps the elements of *Ar* to *V*, indicating a specific value related to the argument; $\succ_{\alpha}$ is a transitive, asymmetric, and irreflexive

relation that shows which values in V the audience α prefers. For example, we assume there are two opposing arguments regarding media censorship.

Argument A: Media should be censored for the safety of the public due to the prevalence of misinformation.

Argument B: Media censorship prevents freedom of the press.

These two arguments constitute *Ar*. While argument A supports media censorship, argument B opposes it. Therefore, *att* denotes that the two arguments are in an attacking relationship. *V* comprises the value of public security and freedom of the press. *val* represents how the value of public security is related to argument A and the value of press freedom is related to argument B. The value in question for argument A is denoted as *val(a)* and *val(b)* for argument B. Audience α's preference for valuing the safety of the public over freedom of the press can be expressed as the following:

$$val(a) \succ_{\alpha} val(b)$$

As audiences have different value-preference profiles ($\succ_{\alpha}$), the evaluation of the argument is subjective. According to Atkinson and Bench-Capon (2021), argument A has more rhetorical quality than argument B for audience α when the value that of argument A is preferred over that of argument B: $val(a) \succ_{\alpha} val(b)$. For audience β, argument B may be more convincing than argument A if β considers *val(b)* to be more important than *val(a)*. An argument can only be considered to be *objectively* persuasive or of higher rhetorical quality when the value the argument promotes is the most preferred value across all audiences (Atkinson & Bench-Capon, 2021).

The core mechanisms behind the subjectivity of rhetorical quality are aligned with cognitive dissonance theory, prior belief effect, matching effect, and more. Festinger's cognitive dissonance theory (Festinger, 1957) suggests that people have an inner drive to hold their attitudes, beliefs, and behaviors in harmony and avoid disharmony (i.e., dissonance). According to the belief disconfirmation paradigm of cognitive dissonance theory (Harmon-Jones, 2002), people experience dissonance when faced with information that is inconsistent with their beliefs. Out of the various ways to reduce dissonance (e.g., changing beliefs and misinterpreting the information), the mechanism of AVAF is aligned with people rejecting and refuting the information to preserve self-consistency. Additionally, the prior belief effect suggests that arguments that are inconsistent with prior beliefs are subject to more extensive refutation and are judged to be weaker (Edwards & Smith, 1996; Weber et al., 2015). People are biased against arguments that are incompatible with their existing opinions (Lord et al., 1979). Similarly, the matching effect states that messages with features that are congruent with a given audience characteristic are likely to have a stronger effect on the audience's belief and behavior (Rothman et al., 2020). Therefore, the individual differences in the evaluation of rhetorical quality can be influenced by the audience's prior beliefs, personal preferences, backgrounds, and general value systems, such as moral values (Batson, 1975; Chapman & Chapman, 1959; Darley & Gross, 1983; Habernal & Gurevych, 2016; Luttrell et al., 2017).

Moral value-system

According to AVAF, constructing the value-preference profiles of argument receivers or audiences provides insight into understanding and predicting argument evaluation. As mentioned above, this study focuses on a value system with general relevance for a multitude of arguments and debates. Specifically, we focus on the moral framing of arguments and the corresponding moral value system of the audience. Moral values and judgments are the underlying fundamentals of people's attitudes and identities (Aquino & Reed, 2002; Strohminger & Nichols, 2014). Therefore, individual differences in argument evaluation (at least partially) can be attributed to audiences' idiosyncratic moral values. People tend to find arguments particularly persuasive when they appeal to moral values that align with their own, deeply held moral beliefs

(e.g., Aramovich et al., 2012; Feinberg & Willer, 2015; Koleva et al., 2012; Luttrell et al., 2017; Ryan, 2017; Weber et al., 2015). It has also been argued that the arguments people communicate are "inevitably moral inducements" (Fisher, 1984, pg. 2) and people construct arguments based on their moral intuitions (Feinberg & Willer, 2019). For instance, a person who believes that it is morally correct to respect a rightful authority (e.g., Centers for Disease Control and Prevention) may construct an argument that advocates a mask mandate. Someone who emphasizes personal freedom in his or her moral value system is likely to discredit the same argument as it conflicts with their core moral beliefs. In addition, an argument that people do not initially support can be made appealing by emphasizing the moral values that are salient to them (see moral reframing in Feinberg & Willer, 2019). Therefore, arguments that are framed in terms of moral values and are personally relevant to the audience are likely to be more persuasive (Jensen et al., 2012).

The content-receiver similarity further explains how an individual's moral intuitions affect their evaluation of arguments with specific moral content. It is assumed that the higher the similarity between the moral content of an argument and the moral values of an audience is, the higher is the rhetorical quality and persuasiveness of that particular argument (Simons et al., 1970). According to the homophily principle, people prefer others who are similar to them (Rogers & Bhowmik, 1970). As the moral content of an argument is reflective of the moral values of the argument producers, the similarity between the moral content of the argument and the moral values of the audience can be translated as the similarity between the argument producers and the argument receivers or audience. Those with similar moral profiles view each other as ingroup members who are trustworthy and credible (Cohen, 2003; Kalkhoff & Barnum, 2000).

Moral Foundations Theory (MFT; Haidt, 2007) provides a theoretical frame to construct a profile that highlights the moral values preferred by the audience of the argument (i.e., moral value system profiles) or emphasized in the presented argument (i.e., moral content profiles). According to MFT, all individuals have an innate sense of moral knowledge that stems from recurring social problems and opportunities faced by the species over long periods of time (Haidt, 2007). These foundations represent the building blocks of morality by "combining previous experiences and emotions into intuitive bits of mental structures" (Tamborini, 2011, pg. 40). However, MFT also posits that this innate moral knowledge is not static and becomes modified through cultural learning, allowing specific moral foundations to become more salient as a function of multiple exogenous and endogenous pressures (Graham et al., 2009). While humans can vary in the degree to which they endorse different moral foundations, MFT indicates that five moral domains are present across cultures (i.e., innate and universally-shared moral intuitions): (1) Care and harm are related to the intuitions of compassion, nurturance, and sympathy; (2) fairness and cheating refer to the sense of justice and righteousness; (3) loyalty and betrayal are related to the moral obligation of group spirit and patriotism; (4) authority and subversion refer to the concerns about maintaining social order and obedience; and (5) sanctity and degradation include moral disgust and the spiritual concerns related to someone's body. Although Haidt (2012) suggested a sixth foundation, this study focuses on the five foundations as they have been extensively validated without the addition of the sixth foundation. A moral foundation can be either upheld (i.e. care, fairness, loyalty, authority, and sanctity) or violated (i.e. harm, cheating, betrayal, subversion, and degradation). Thus, audiences' moral value systems and arguments' moral content profiles can be constructed with a 10-dimensional vector (5 foundations $\times$ 2 uphold/violate). Each dimension signifies the salience of a moral foundation to the audience or its representation in the framing of an argument. Given the discussion above, audiences are more likely to find an argument persuasive or convincing if its moral content profile is similar to their moral value system. Additionally, we exploratively investigate how similarity in arguments' moral content profile and audiences' moral value system (i.e., value-driven MAS) interacts with value-free MAS as examined in study 1.

H3: The more the arguments' moral content profile is similar to the audiences' moral value system, the higher the audience evaluates the argument's rhetorical quality.

RQ4: Does the relationship between rhetorical quality of an argument and its (a) information-source citation, (b) sentiment, (c) length, (d) concreteness, and (e) quantitative specificity change depending on the similarity between its moral content profile and the audience's moral value system?

Methods

Study 2 measured the similarity between the moral content profiles of the arguments made in either the pro- or the contra- stance and the moral value system of the audience, and tested its effect on rhetorical quality. The moral profiles were constructed by computationally extracting moral signals using the extensively validated Extended Moral Foundations Dictionary (eMFD; Hopp & Weber, 2021; Hopp et al., 2021) and its scoring tool: eMFDscore.[10] eMFD is built on crowdsourced annotations from a large and highly diverse textual corpus. For each word in the dictionary, eMFD provides five vector probabilities of the word belonging to each of the foundations in MFT, and a score that indicates its valence (positive/upheld or negative/violated). The upholding and violating moral foundations refer to virtue and vice dimensions in eMFDscore. Thus, using eMFDscore results in ten moral foundation probabilities (5 vice and 5 virtue dimensions) for extracted text content. The operationalization of moral profiles relies on the assumption that people who are sensitive to certain moral foundations are likely to use more words related to those foundations (Araque et al., 2022; Lai et al., 2021; Matsuo et al., 2019).

The moral content profiles of the arguments made by the argument producers were obtained by running eMFDscore on the arguments that are aggregated for each stance. The moral value systems of the audience (i.e., the argument receivers who voted on the arguments made by two opposing sides) were constructed by calculating the ten moral foundation probabilities as extracted via eMFDscore on all historical arguments the audience made (i.e., all the arguments made by an audience across multiple debates they have participated in as either the instigator or the contender). There were around a quarter of users without any historical arguments ($n = 15{,}629$, 25.28%): these users are omitted in the analysis of study 2. For the users with historical arguments, on average, they participated in 33 debates ($min = 10$, $max = 661$, Q1 = 13, Q2 = 20, Q3 = 35, $SD = 47.56$).

The similarity in moral profiles was calculated by computing the difference in cosine similarity between the moral profile of the pro-stance argument and the audience, and between that of the contra-stance argument and the audience. The similarity measure is adopted in this study as an operationalization of value-driven MAS as defined above. A positive value indicates that the moral profile between the pro-stance argument and the audience was more similar than between the contra-stance argument and the audience. The value-driven MAS ranged from −1.00 to 1.00 (Q1 = −0.02, Q2 = 0, Q3 = 0.02) with a mean of −0.003 ($SD = 0.12$).

Results

Study 2 also uses the data used in study 1 with another level of data screening. We conducted a mixed-effects binomial logistic regression after considering only the votes made by users with historical debates and debates with more than 2 votes. For study 2, we analyze a total of 27,534 rhetorical quality judgments on 5,990 debates. We constructed an unconditional mean model (see Model 3 in Table 1) and a model with value-free MAS as fixed effects (see Model 4 in Table 1) to examine if this smaller sample of rhetorical quality judgments changed the results. Multicollinearity was not a problem in all the models examined in study 2: the largest VIF was 1.81 in Model 6.

[10] https://github.com/medianeuroscience/emfdscore

Table 2. Summary of mixed-effects binomial logistic regression for value-driven MAS (Study 2).

Fixed Effects	Study 2: Model 5	Study 2: Model 6
(Intercept)	0.55***[0.51–0.58]	0.54***[0.51–0.58]
Information-source citation	1.49***[1.38–1.59]	1.49***[1.39–1.60]
Quantitative specificity	0.66***[0.55–0.80]	0.67***[0.56–0.82]
Length	4.41***[4.00–4.85]	4.38***[3.98–4.83]
Sentiment	0.92* [0.86–0.98]	0.92* [0.87–0.98]
Concreteness	0.74***[0.69–0.79]	0.71***[0.66–0.76]
Similarity in moral profile (SMP)	1.28***[1.19–1.37]	1.26***[1.17–1.35]
SMP × Information-source citation		1.07 [0.95–1.20]
SMP × Quantitative specificity		0.92 [0.75–1.12]
SMP × Length		1.07 [0.95–1.20]
SMP × Sentiment		1.05 [0.98–1.14]
SMP × Concreteness		1.02 [1.00–1.03]
Random Effects		
σ^2	3.29	3.29
τ_{00}	3.95	3.94
ICC	0.55	0.55
Marginal R^2	0.28	0.28
Conditional R^2	0.67	0.67
Number of debates (*k*)	5,990	5,990
Number of votes (*n*)	27,534	27,534

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. For the fixed effects, the odds ratios are provided and the 95% confidence intervals using a Wald z-distribution approximation are provided in brackets. All predictors are standardized.

Model 3 shows that 65% of the variance in rhetorical quality can be explained by the debate level grouping. The addition of value-free MAS, significantly improved the model, X^2 (5) = 2105.6, $p < .001$. The fixed effects alone explain 27% of the variance in rhetorical quality. A stance with a persuasive argument is characterized by having more information-source citation, less quantitative specificity, more unique words, more negative sentiment, and more abstract language. Unlike study 1, the effect of sentiment was significant (odds ratio = 0.93). When considering the votes made by those who previously have engaged in the debate platform as argument producers, providing more negative sentiment by one standard deviation more than the opposing side (i.e., 0.93 more) increases the odds of being identified as the more convincing side by 7%.

To examine the effect of value-driven MAS, the similarity in the moral profile measure was added as a fixed effect (see Model 5 in Table 2), which significantly improved the model, X^2 (1) = 50.7, $p < .001$. The value-driven MAS explained 1% more variance in rhetorical quality above and beyond value-free MAS. When a stance of a debate has more similar moral content profile with the moral profile of the argument receiver by one standard deviation (i.e., 0.12 higher in similarity measure), the argument receiver is 28% more likely to consider this stance as more persuasive.

To explore the interaction between value-free MAS and value-driven MAS, we added their product terms to the model as fixed effects (see Model 6 in Table 2). Although the model with five additional product terms is statistically different from the previous model [X^2 (5) = 11.35, $p = .04$], none of the predictors were statistically significant. The interaction between value-driven MAS and concreteness was the only interaction that was nearly significant with odds ratio of 1.02 ($p = .051$).

Discussion

Study 2 demonstrated that value-driven MAS that focus on the moral framing of the arguments and the moral value systems of argument receivers is a statistically significant predictor of argument quality (although the effect size is small, which is discussed further in the general discussion section). This finding is consistent with the literature: the argument receivers' values and beliefs affect the evaluation of argument quality (Lukin et al., 2017; Sherman & Cohen, 2002). More specifically, the receivers consider the argument that is consistent with their beliefs to be persuasive (J. A. L. Hoeken & van Vugt, 2016; Mercier & Sperber, 2011). This study not only corroborates Wachsmuth, Naderi, Hou,

et al. (2017) claim about how argument quality is subjective depending on the perception of the audience but also demonstrates how the subjectivity can be captured using computationally extracted moral framing and moral value systems using MFT and eMFD.

Inconsistent with study 1, study 2 found that sentiment is significantly related to rhetorical quality. While study 1 included users who are less actively using the platform or have little experience with it, study 2 only examined those who have previous experience in constructing their own argument in the platform. Therefore, the filtered argument audiences are more likely to evaluate arguments more thoroughly and may be more familiar with argumentation, easily noticing the salience of sentiment.

As for the interaction between the value-driven MAS and value-free MAS, the results indicate that the interaction between value-driven MAS and concreteness was close to statistical significance. Although the interpretation should be made with caution, this may imply how the value systems of the audience can affect the way they process arguments. According to Slater and Rouner (1996), argument receivers engage in biased processing when the argument is incongruent with their own values. Considering the findings of study 2 and Menegatti and Rubini's (2013) that found abstract messages to be more effective toward someone with a similar political stance, future research can examine if the effect of abstract language is moderated by only political stance or also by other value systems.

General discussion

The two studies demonstrate how rhetorical quality of an argument can be predicted at scale by using value-free and value-driven MAS. In dialectic argumentation, the side that uses more information-source citation, less quantitative specificity, more unique words, and abstract language is likely to be more persuasive. In study 2, we focused on the effect of moral framing in arguments and the argument receiver's moral value systems (i.e., value-driven MAS). The results indicate that audiences are likely to find the stance that addresses similar moral values to their moral value system as more persuasive. Additionally, study 2 also found a significant effect of sentiment on rhetorical quality but no statistically significant interaction between value-free and value-driven MAS.

Although the six measures of argument strength proposed in this study were statistically significant, more than 50% of the variance in rhetorical quality was explained by the debate-level grouping. This may reflect the social aspect of the debate platform: users can see other users' votes. Consequently, the voting behavior can be subject to the spiral of silence: people are less willing to express their opinion when it is inconsistent with the public opinion (Noelle-Neumann, 1974). Users are more likely to vote when their rhetorical judgment is aligned with other voters. It should be noted that the value-free and value-driven MAS explained 28% of the variance of rhetorical quality (without the random effect of the debate level groupings).

Compared to value-free MAS, the addition of value-driven MAS via the similarity of arguments' moral framing and the audience's moral value system explained around 1% of the variance in rhetorical quality above and beyond the value-free MAS. The small effect size can be explained from theoretical, methodological, and empirical perspectives. The audience's value system can be conceptualized at various levels. Although MFT states that the examined moral values are foundational, they may be too rudimentary and abstract. Constructing people value systems based on social issues, such as abortion, immigration, and gun control, may enhance the predictive utility of value-driven MAS. Additionally, some debates were less moralized. There were more than 4,000 debates that were categorized as travel, cars, fashion, TV, movies, music, arts, funny, games, technology, and sports by *Debate.org*. From a methodological perspective, eMFD and its scoring method can be further improved. Previous studies that used eMFD or other moral dictionaries have also found small effect sizes, explaining less than 3% of variance of the dependent variable (e.g., Hopp et al., 2021; Matsuo et al., 2019; Rezapour et al., 2021). Therefore, the effect of value-driven MAS can increase with the advancement of computational approaches to extracting moral content. On an empirical ground, Rains et al. (2018) analysis of

60 years of quantitative communication research shows that the median effect size of persuasion studies is $r = .13$ ($SD = 0.19$), which is around 1.7% of explained variance. To put it differently, the effect of value-driven MAS is small but falls within the acceptable within the persuasion literature.

Durmus and Cardie's (2019) on *Debate.org* has also examined the audience's value system, but they relied on self-reports. They found that prior belief (i.e., religion or political ideology) played a larger role than linguistic features in predicting argument strength, which is inconsistent with our findings. The difference in results may be due to the following reasons: (1) Durmus and Cardie only examined debates that were relevant to the prior belief (i.e., debates in religion category are examined when using religious ideology to predict argument quality), (2) they operationalized prior belief as a dichotomous variable (i.e., liberal vs. conservative; atheist vs. christian), and (3) their dependent variable is an aggregate of various argument evaluations (e.g., convincingness, spelling, and use of reliable sources), which obscures the conceptual definition of argument strength. Although both research examined *Debate.org*, this research provides additional contributions to the argumentation and persuasion literature as (1) theory driven value-free MAS are examined (although not exhaustive); (2) the operationalization of argument quality is specific and consistent with the literature; (3) demonstration is provided that value-free and value-driven MAS can feasibly be applied across various debate topics. However, the predictive power of value-driven MAS above and beyond the value-free MAS is small, and the practical implications should be considered with caution.

Developing a better understanding of the characteristics that drive argument quality unrelated to self-reports, in naturalistic settings, at scale is instrumental for several reasons. First, by examining MAS, online discussions are beneficial for advancing theories and methods of argumentation, persuasion, and attitude-change research that have largely been tested in controlled experimental settings (Zhao & Cappella, 2016; Zhao et al., 2011). An online debate forum provides a valuable naturalistic dataset to study argumentation and persuasion as the platform focuses on the construction of debates rather than reacting to a stance as users do in social media (Dutta et al., 2020). As there are various debate topics that are available, online debate forums can also contribute to moral persuasion research, which usually focuses on one or two topics (e.g., Kodapanakkal et al., 2022).

Persuasion research that has followed survey logic in constructing arguments considers an argument to be persuasive because it intuitively seems persuasive. For instance, researchers used the argument that participants from a pilot study indicated as having a higher quality or soliciting a more persuasive outcome as the persuasive argument in the main experiment (O'Keefe & Jackson, 1995). This tautological operationalization of argument quality does not explain why an argument is persuasive (Hahn, 2020; Petty & Cacioppo, 1986). Although further validations and replications are required, the present research demonstrated how value-free and value-driven MAS can be computationally extracted. MAS can be used instead of tautologically reasoned self-reports as a manipulation check for argument quality in experiments.

There has been a recent increase in the popularity of online discussion forums, which have the potential to shape public opinion and trends in online polarization (e.g., Wang et al., 2018). Considering how moral arguments strengthen people's moral convictions (Kodapanakkal et al., 2022), moralized debates may facilitate polarization and moral divides (Aramovich et al., 2012; Ryan, 2017). Hence, examining arguments' strengths and weaknesses that permeate individuals' argumentation and reasoning using automated, computational algorithms is crucial for both debate platform designers, users, and those with an interest in influencing public opinion. By identifying common patterns that underlie argumentation and argument appraisal, platform developers can implement better algorithms that assist users in the argument construction process and thereby improve overall argument and discussion quality. Furthermore, pointing users to the composition of a weak argument may improve argumentation beyond online spheres and lead to more deliberative everyday discussions that are foundational for a functioning democracy (Sinnott-Armstrong, 2018). Analogously, the rise of misinformation and fake news has further emphasized the need for educating citizens in carefully scrutinizing the semantic content, veracity, and logical structure of arguments.

Although the applicability of the MAS to other online platforms, such as social media, remains untested, the framework of computationally extracting value-free and value-driven MAS may be used to enhance content monitoring, moderation, and recommendation. The ongoing contention surrounding content moderation revolves around who should be responsible for moderation and what should be moderated (Samples, 2019). However, the strength of arguments presented in the content is often overlooked. Given the variability of argument persuasiveness, content monitoring should not solely concentrate on its subject matter, such as terrorism and fake news, but also prioritize a monitoring protocol based on the content's persuasive features. Furthermore, Colleoni et al. (2014) examined echo-chambers on Twitter by predicting political orientation based on users' shared content. Building upon this type of research, value-driven MAS can be used to investigate how an online platform promotes echo-chambers by differentiating the effect of the argument topic, value-free, and value-driven MAS.

There are several limitations that must be taken into account when interpreting the findings. First, caution is needed when generalizing the results to other debate platforms beyond the one examined in this study. Despite the diversity in the debate topics and the large number of rhetorical quality judgments, we focused on a single platform, and the results depend on its specific structure and features. Therefore, our conclusion that persuasiveness is influenced by factors such as citation, unique words, quantitative specificity, and abstract words mainly applies to dialectical argumentation where two opposing sides are present, and the audience evaluates which side makes a more convincing argument. In other debate platforms, such as *Kialo.com*, where multiple users exchange short arguments, audiences evaluate each argument rather than the overall stance. Consequently, the current measures of value-free and value-driven MAS may be most useful for comparing two opposing arguments directly. Additionally, generalizing the results to other social media platforms, such as Twitter, should also be done with caution. Although social media users can also exchange arguments, the affordance of the platform, such as availability of images and limitations in word count, as well as the familiarity of the user make other factors of persuasion (that are not examined in this study) relevant. For instance, the authority of the user on social media may play a role in persuasive discourse, but not in an online debate platform where the users are relatively more anonymous.

Second, we purposefully operationalized MAS using a simple string matching function in Python and a dictionary-based approach to ensure accessibility of the measures and interpretability of the results for a larger body of communication scholars. Although the operationalization is clear for some MAS (e.g., quantitative specificity), there is room for improvement in others. For instance, measuring information-source citation using hyperlinks requires the platform to predominantly use hyperlinks, which was the case for Debate.org. Additionally, a hyperlinked source can be a news article, a research paper, or even a Wikipedia page. Considering how these sources vary in the reliability of their information and fact-checking procedures, integrating the quality of the hyperlink may further enhance the predictability of argument quality. Although VADER uses a rule-based model for calculating the sentiment, outperforms other sentiment dictionaries, and takes 90% of negations into account (Bonta et al., 2019; Hutto & Gilbert, 2014), there is evidence for deep learning models trained on a specific dataset (e.g., COVID-19 tweets) outperforming VADER (Rustam et al., 2021). As for measuring concreteness and constructing moral profiles, more advancement and scholarly attention is needed as a validated alternative to the dictionary-based approach and is currently not available or is in development (Atari et al., 2023; Solovyev, 2021). We recommend future researchers to weigh the costs (e.g., ignoring semantics and contexts) and benefits (e.g., accessible and interpretable) of using validated dictionary-based approaches as opposed to creating or using alternative methods.

Third, there are other potential MAS, such as clarity and comprehensiveness (B. T. Johnson et al., 2005). Integrating text-analysis APIs that measure reliability (e.g., readable.com and translatedlabs.com/text-readability) can further expand the list of value-free MAS. Last, Blair and Johnson (1987) listed the necessary conditions for a reliable and objective evaluation of argumentation: the evaluators need to be knowledgeable about the topic of argumentation, be reflective, persistently ask questions, and be open to change in opinion. This research did not

examine to what degree these conditions were met. Future research should develop computational methods to examine these qualities of the audiences. For instance, intellectual humility, which refers to being open, engaging, modest, and corrigible to new ideas and knowledge (see Alfano et al., 2017 for more), can be measured using a self-report survey. However, a valid text-driven computational approach to measure intellectual humility is yet to be developed and implemented to further advance our understanding of persuasion.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Sungbin is a Ph.D. candidate in the Department of Communication at the University of California, Santa Barbara, and a researcher in the Media Neuroscience Lab. His research focuses on the accurate assessment and effective resolution of problematic media use. His research integrates (1) computational and neuroscience approach to persuasion, and (2) psychological and behavioral approach to excessive media consumption, such as short-form videos. By bridging these interdisciplinary perspectives, he aims to provide a comprehensive understanding of problematic media use and contribute to the development of evidence-based strategies to promote healthier media habits.

Musa Malik is a graduate student in the Department of Communication at the University of California, Santa Barbara, and a researcher in the Media Neuroscience Lab. He holds a BS in Neuroscience from New York University, Shanghai and a MA in Communication from the University of California, Santa Barbara.

Yibei Chen received her Ph.D. in Communication from the University of California, Santa Barbara, and is an incoming Postdoctoral Associate at the McGovern Institute for Brain Research at Massachusetts Institute of Technology (MIT). Her research focuses on narrative comprehension (i.e., information processing) in the brain and mind.

Frederic R. Hopp (Ph.D., UC Santa Barbara) is Assistant Professor for Political Communication at the University of Amsterdam's School of Communication Research. He is interested in the moral content of human communication and how moralized messages are cognitively processed and motivate behavior. Frederic leverages a method-theory synergy that combines behavioral experiments, neuroimaging, natural language processing, and machine learning. His work has received numerous awards and been published in scientific journals including Nature Human Behavior, Journal of Communication, Behavior Research Methods, and Computational Communication Research.

René Weber received his Ph.D. (Dr.rer.nat.) in Psychology from the University ofTechnology in Berlin, Germany, and his M.D. (Dr.rer.medic.) in Psychiatry and CognitiveNeuroscience from the RWTH University in Aachen, Germany. He is a Professor in theDepartment of Communication and the Department of Psychological and Brain Sciencesat the University of California in Santa Barbara, director of UCSB's Media NeuroscienceLab (https://medianeuroscience.org), and member of UCSB's Neuroscience Institute(https://www.nri.ucsb.edu). He also holds a Visiting Professorship position at EwhaWomans University in Seoul, South Korea. He was among the first media psychologyscholars who regularly use computational approaches and brain imaging technology toinvestigate various topics related to media industries, from the appeal of mediaentertainment, diversity and inclusion in the media, the impact of media violence, to thepersuasiveness of campaigns. He has published four books and more than 150 journalarticles and book chapters. His research has been supported by grants fromnational scientific foundations in the United States and Germany, as well as throughprivate philanthropies and industry contracts. He is a Fellow of the InternationalCommunication Association.

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., Lapsley, D., & Tractenberg, R. E. Development and validation of a multi-dimensional measure of intellectual humility. (2017). *PloS One*, *12*(8), e0182950. https://doi.org/10.1371/journal.pone.0182950

Aquino, K., & Reed, I. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83* (6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Aramovich, N. P., Lytle, B. L., & Skitka, L. J. (2012). Opposing torture: Moral conviction and resistance to majority influence. *Social Influence*, *7*(1), 21–34. https://doi.org/10.1080/15534510.2011.640199

Araque, O., Gatti, L., & Kalimeri, K. (2022). LibertyMFD: A Lexicon to assess the moral foundation of liberty. *Proceedings of the Conference on Information Technology for Social Good*, *2*, 154–160. https://doi.org/10.1145/3524458.3547264

Atari, M., Omrani, A., & Dehghani, M. (2023). Contextualized construct representation: Leveraging psychometric scales to advance theory-driven text analysis. *PsyArxiv*. https://doi.org/10.31234/osf.io/m93pd
Atkinson, K., & Bench-Capon, T. (2021). Value-based Argumentation. *Journal of Applied Logics -IfColog Journal of Logics and Their Application*, *8*(6), 1543–1588. http://collegepublications.co.uk/ifcolog/?00048
Baesler, E. J., & Burgoon, J. K. (1994). The temporal effects of story and statistical evidence on belief change. *Communication Research*, *21*(5), 582–602. https://doi.org/10.1177/009365094021005002
Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on stated religious belief. *Journal of Personality and Social Psychology*, *32*(1), 176–184. https://doi.org/10.1037/h0076771
Bell, B. A., Ferron, J. M., & Kromrey, J. D. (2008). Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models. *JSM Proceedings: Section on Survey Research Methods*, 1122–1129. http://www.asasrms.org/Proceedings/y2008f.html
Bench-Capon, T. J. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, *13*(3), 429–448. https://doi.org/10.1093/logcom/13.3.429
Blair, J. A. (2012). Argumentation as rational persuasion. *Argumentation*, *26*(1), 71–81. https://doi.org/10.1007/s10503-011-9235-6
Blair, J. A., & Johnson, F. H. (1987). Argumentation as dialectical. *Argumentation*, *1*(1), 41–56. https://doi.org/10.1007/BF00127118
Bonta, V., Janardhan, N. K. N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037
Boote, A. S. (1981). Market segmentation by personal values and salient product attributes. *Journal of Advertising Research*, *21*(1), 29–35. https://psycnet.apa.org/record/1981-22299-001
Borah, P. (2014). The hyperlinked world: A look at how the interactions of news frames and hyperlinks influence news credibility and willingness to seek information. *Journal of Computer-Mediated Communication*, *19*(3), 576–590. https://doi.org/10.1111/jcc4.12060
Brosius, H. B. (2000). Toward an exemplification theory of news effects. *Document Design*, *2*(1), 18–27. https://doi.org/10.1075/dd.2.1.03bro
Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5
Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*(1), 3–25. https://doi.org/10.1207/s15327957pspr0101_2
Carpenter, C. J. (2015). A meta-analysis of the ELM's argument quality× processing type predictions. *Human Communication Research*, *41*(4), 501–534. https://doi.org/10.1111/hcre.12054
Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, *58*(3), 220–226. https://doi.org/10.1037/h0041961
Cialdini, R. B. (2008a). *Influence. Science and practice*. Allyn and Bacon.
Cialdini, R. B. (2008b). Turning persuasion from an art into a science. In P. Meusburger, M. Welker, & E. Wunder (Eds.), *Clashes of knowledge* (pp. 199–209). Springer Netherlands.
Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health*, *62*(8), 752–758. https://doi.org/10.1136/jech.2007.060798
Clarke, P., & Wheaton, B. (2007). Addressing data sparseness in contextual population research using cluster analysis to create synthetic neighborhoods. *Sociological Methods & Research*, *35*(3), 311–351. https://doi.org/10.1177/0049124106292362
Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822. https://doi.org/10.1037/0022-3514.85.5.808
Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317–332. https://doi.org/10.1111/jcom.12084
Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33. https://doi.org/10.1037/0022-3514.44.1.20
Dillard, J. P., & Pfau, M. (2002). *The persuasion handbook: Developments in theory and practice*. Sage Publications. https://doi.org/10.4135/9781412976046
Durmus, E., & Cardie, C. (2019). Exploring the role of prior belief for argument persuasion. arXiv. https://doi.org/10.48550/arXiv.1906.11301
Dutta, S., Das, D., & Chakraborty, T. Changing views: Persuasion modeling and argument extraction from online discussions. (2020). *Information Processing & Management*, *57*(2), 102085. https://doi.org/10.1016/j.ipm.2019.102085
Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*(1), 5–24. https://doi.org/10.1037/0022-3514.71.1.5
Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, *41*(12), 1665–1681. https://doi.org/10.1177/0146167215607842

Feinberg, M., & Willer, R. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, *13*(12). https://doi.org/10.1111/spc3.12501.
Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
Fisher, W. R. (1984). Narration as a human communication paradigm: The case of public moral argument. *Communication Monographs*, *51*(1), 1–22. https://doi.org/10.1080/03637758409390180
Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*(6), 889–906. https://doi.org/10.1037/0022-3514.38.6.889
Gilbert, M. A. (2004). Emotion, argumentation, and informal logic. *Informal Logic*, *24*(3), 245–264. https://doi.org/10.22329/il.v24i3.2147
Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029. https://doi.org/10.1037/a0015141
Habernal, I., & Gurevych, I. (2016). Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM. In A. van den Bosch (Ed.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp.1589–1599). Association for Computational Linguistics. https://aclanthology.org/P16-1150/
Hahn, U. (2020). Argument quality in real world argumentation. *Trends in Cognitive Science*, *24*(5), 363–374. https://doi.org/10.1016/j.tics.2020.01.004
Haidt, J. (2007). The new synthesis in moral psychology. *Science: Advanced Materials and Devices*, *316*(5827), 998–1002. https://doi.org/10.1126/science.1137651
Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
Hansen, J., & Wänke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality and Social Psychology Bulletin*, *36*(11), 1576–1588. https://doi.org/10.1177/0146167210386238
Harmon-Jones, C. (2002). A cognitive dissonance theory perspective on persuasion. In J. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 99–116). Sage Publications.
Heit, E., & Rotello, C. M. (2012). The pervasive effects of argument length on inductive reasoning. *Thinking & Reasoning*, *18*(3), 244–277. https://doi.org/10.1080/13546783.2012.695161
Hitchcock, D. (2006). Informal logic and the concept of argument. In D. Jacquette (Ed.), *Philosophy of logic* (pp. 101–129). Elsevier.
Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation*, *15*(4), 425–437. https://doi.org/10.1023/A:1012075630523
Hoeken, H., & Hustinx, L. (2009). When is statistical evidence superior to anecdotal evidence in supporting probability claims? The role of argument type. *Human Communication Research*, *35*(4), 491–510. https://doi.org/10.1111/j.1468-2958.2009.01360.x
Hoeken, H., Timmers, R., & Schellens, P. J. (2012). Arguing about desirable consequences: What constitutes a convincing argument? *Thinking & Reasoning*, *18*(3), 394–416. https://doi.org/10.1080/13546783.2012.669986
Hoeken, J. A. L., & van Vugt, M. (2016). The biased use of argument evaluation criteria in motivated reasoning: Does argument quality depend on the evaluators' standpoint? In F. Paglieri, L. Bonelli, & S. Felletti (Eds.), *The psychology of argument: Cognitive approaches to argumentation and persuasion* (pp. 197–210). College Publications.
Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, *53*(1), 232–246. https://doi.org/10.3758/s13428-020-01433-0
Hopp, F. R., & Weber, R. (2021). Reflections on extracting moral foundations from media content. *Communication Monographs*, *88*(3), 371–379. https://doi.org/10.1080/03637751.2021.1963513
Hornikx, J., & Hoeken, H. (2007). Cultural differences in the persuasiveness of evidence types and evidence quality. *Communication Monographs*, *74*(4), 443–463. https://doi.org/10.1080/03637750701716578
Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web & Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550
Jensen, J. D., King, A. J., Carcioppolo, N., & Davis, L. (2012). Why are tailored messages more effective? A multiple mediation analysis of a breast cancer screening intervention. *Journal of Communication*, *62*(5), 851–868. https://doi.org/10.1111/j.1460-2466.2012.01668.x
Johnson, B. T., Maio, G. R., & Smith McLallen, A. (2005). Communication and attitude change: Causes, processes, and effects. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 617–669). Lawrence Erlbaum Associates Publishers.
Johnson, K. A., & Wiedenbeck, S. (2009). Enhancing perceived credibility of citizen journalism websites. *Journalism & Mass Communication Quarterly*, *86*(2), 332–348. https://doi.org/10.1177/107769900908600205
Kalkhoff, W., & Barnum, C. (2000). The effects of status-organizing and social identity processes on patterns of social influence. *Social Psychology Quarterly*, *63*(2), 95–115. https://doi.org/10.2307/2695886
Kim, C. (1972). Can men find the meaning of "meaning?. *ETC: Areview of general semantics*, *29*(3), 251–255. https://www.jstor.org/stable/42576447

Kodapanakkal, R. I., Brandt, M. J., Kogler, C., & van Beest, I. (2022). Moral frames are persuasive and moralize attitudes; nonmoral frames are persuasive and de-moralize attitudes. *Psychological Science*, *33*(3), 433–449. https://doi.org/10.1177/09567976211040803

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of Research in Personality*, *46*(2), 184–194. https://doi.org/10.1016/j.jrp.2012.01.006

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodal. *Psychological Inquiry*, *10*(2), 83–109. https://doi.org/10.1207/S15327965PL100201

Lai, M., Stranisci, M. A., Bosco, C., Damiano, R., & Patti, V. (2021). HaMor at the profiling hate speech spreaders on Twitter. *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum*, *2936*, 2047–2055. http://ceur-ws.org/Vol-2936/paper-178.pdf

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37–51). Harper and Row.

Li, C. Y. (2013). Persuasive messages on information system acceptance: A theoretical extension of elaboration likelihood model and social influence theory. *Computers in Human Behavior*, *29*(1), 264–275. https://doi.org/10.1016/j.chb.2012.09.003

Li, H., Liu, H., & Zhang, Z. (2020). Online persuasion of review emotional intensity: A text mining analysis of restaurant reviews. *International Journal of Hospitality Management*, *89*, 102558. Article 102558. https://doi.org/10.1016/j.ijhm.2020.102558

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Lukin, S. M., Anand, P., Walker, M., & Whittaker, S. (2017). Argument strength is in the eye of the beholder: Audience effect in persuasion. arXiv. https://doi.org/10.48550/arXiv.1708.09085

Luttrell, A., Petty, R. E., & Xu, M. (2017). Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology*, *69*, 178–183. https://doi.org/10.1016/j.jesp.2016.09.006

Luu, K., Tan, C., & Smith, N. A. (2019). Measuring online debaters' persuasive skill from text over time. *Transactions of the Association for Computational Linguistics*, *7*, 537–550. https://doi.org/10.1162/tacl_a_00281

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Matsuo, A., Sasahara, K., Taguchi, Y., Karasawa, M., & Gruebner, O. (2019). Development and validation of the Japanese moral foundations dictionary. *PLoS One*, *14*(3), article e0213343. https://doi.org/10.1371/journal.pone.0213343

McCormack, K. C. (2014). Ethos, pathos, and logos: The benefits of aristotelian rhetoric in the courtroom. *Washington University Jurisprudence Review*, *7*(1), 131–155. https://heinonline.org/HOL/P?h=hein.journals/wujurisre7&i=136

Menegatti, M., & Rubini, M. (2013). Convincing similar and dissimilar others: The power of language abstraction in political communication. *Personality and Social Psychology Bulletin*, *39*(5), 596–607. https://doi.org/10.1177/0146167213479404

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M., & Potts, K. A. (2007). Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom. *Human Communication Research*, *33*(2), 219–240. https://doi.org/10.1111/j.1468-2958.2007.00297.x

Nabi, R. L. (1999). A cognitive-functional model for the effects of discrete negative emotions on information processing, attitude change, and recall. *Communication Theory*, *9*(3), 292–320. https://doi.org/10.1111/j.1468-2885.1999.tb00172.x

Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

O'Keefe, D. J. (1995). Argumentation studies and dual-process models of persuasion. In F. van Eemeren, R. Grootendorst, J. Blair, & C. Willard (Eds.), *Proceedings of the Third ISSA Conference on Argumentation*, Amsterdam, The Netherlands (pp.3–17). SIC SAT.

O'Keefe, D. J. (1998). Justification explicitness and persuasive effect: A meta-analytic review of the effects of varying support articulation in persuasive messages. *Argumentation & Advocacy*, *35*(2), 61–75. https://doi.org/10.1080/00028533.1998.11951621

O'Keefe, D. J. (2002). *Persuasion: Theory & research*. Sage Publications.

O'Keefe, D. J., & Jackson, S. (1995). Argument quality and persuasive effects: A review of current approaches. In S. Jackson (Ed.), *Argumentation and values: Proceeding of the 9th SCA/AFA Conference on Argumentation*, Annandale, Virginia, United States (pp. 88–92). Speech Communication Association.

Park, H. S., Levine, T. R., Kingsley Westerman, C. Y., Orfgen, T., & Foregger, S. (2007). The effects of argument quality and involvement type on attitude formation and attitude change: A test of dual-process and social judgment predictions. *Human Communication Research*, *33*(1), 81–102. https://doi.org/10.1111/j.1468-2958.2007.00290.x

Park, H., & Thelwall, M. (2003). Hyperlink analysis of the world wide web: *A review*. *Journal of Computer-Mediated Communication*, *8*(4). https://doi.org/10.1111/j.1083-6101.2003.tb00223.x

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-4964-1

Pierro, A., Mannetti, L., Erb, H.-P., Spiegel, S., & Kruglanski, A. Q. (2005). Informational length and order of presentation as determinants of persuasion. *Journal of Experimental Social Psychology*, *41*(5), 458–469. https://doi.org/10.1016/j.jesp.2004.09.003

Priniski, J., & Horne, Z. (2018). Attitude Change on Reddit's Change My View. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, *40*, 2276–2281. https://cogsci.mindmodeling.org/2018/papers/0437/0437.pdf

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Psychological and Cognitive Sciences*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Rains, S. A., Levine, T. R., & Weber, R. (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association*, *42*(2), 105–124. https://doi.org/10.1080/23808985.2018.1446350

Rezapour, R., Dinh, L., & Diesner, J. (2021). Incorporating the measurement of moral foundations theory into analyzing stances on controversial topics. In O. Conloan & E. Herder (Eds.). *Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (pp. 177–188). Association for Computing Machinery. https://doi.org/10.1145/3465336.3475112

Rogers, E. M., & Bhowmik, D. K. (1970). Homophily-heterophily: Relational concepts for communication research. *The Public Opinion Quarterly*, *34*(4), 523–538. https://doi.org/10.1086/267838

Rothman, A. J., Desmarais, K. J., & Lenne, R. L. (2020). Moving from research on message framing to principles of message matching: The use of gain-and loss-framed messages to promote healthy behavior. *Advances in Motivation Science*, *7*, 43–73. https://doi.org/10.1016/bs.adms.2019.03.001

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., Choi, G. S., & Mumtaz, W. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PloS One*, *16*(2), article e0245909. https://doi.org/10.1371/journal.pone.0245909

Ryan, T. J. (2017). No compromise: Political consequences of moralized attitudes. *American Journal of Political Science*, *61*(2), 409–423. https://doi.org/10.1111/ajps.12248

Sadoski, M., Goetz, E. T., & Rodriguez, M. (2000). Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text types. *Journal of Educational Psychology*, *92*(1), 85–95. https://doi.org/10.1037/0022-0663.92.1.85

Samples, J. (2019). Why the government should not regulate content moderation of social media. *Cato Institute Policy Analysis*, 865. https://ssrn.com/abstract=3502843

Shen, F., & Edwards, H. H. (2005). Economic individualism, humanitarianism, and welfare reform: A value-based account of framing effects. *Journal of Communication*, *55*(4), 795–809. https://doi.org/10.1111/j.1460-2466.2005.tb03023.x

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119–123. https://doi.org/10.1111/1467-8721.00182

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*(1), 1–16. https://doi.org/10.1037/h0028429

Sinnott-Armstrong, W. (2018). *Think again. How to reason and argue*. Oxford University Press.

Slater, M. D., & Rouner, D. (1996). Value-affirmative and value-protective processing of alcohol education messages that include statistical evidence or anecdotes. *Communication Research*, *23*(2), 210–235. https://doi.org/10.1177/009365096023002003

Solovyev, V. (2021). Concreteness/Abstractness concept: State of the art. In B. M. Velichkovsky, P. M. Balaban, V. L. Ushakov, & L. V (Eds.), *Advances in cognitive research, artificial intelligence and neuroinformatics* (pp. 275–283). Springer. https://doi.org/10.1007/978-3-030-71637-0_33

Stavraki, M., Lamprinakos, G., Briñol, P., Petty, R. E., Karantinou, K., & Díaz, D. (2021). The influence of emotions on information processing and persuasion: A differential appraisals perspective. *Journal of Experimental Social Psychology*, *93*, 104085. https://doi.org/10.1016/j.jesp.2020.104085

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology: Theories, Methods, & Applications*, *23*(1), 39–45. https://doi.org/10.1027/1864-1105/a000031

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussion. In J. Bourdeau, J. A. Hendler, & R. N. Nkambou (Eds.), *WWW' 16:Proceedings of the 25th International Conference on the World Wide Web* (pp. 613–624). IW3C3. https://doi.org/10.1145/2872427.2883081

van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E. C. W., Plantin, C., & Walton, D. N. (2009). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge.

van Eemeren, F. H., Grootendorst, R., Jackson, S., & Jacobs, S. (1993). *Reconstructing argumentative discourse*. University of Alabama Press.

Villata, S., Cabrio, E., Jraidi, I., Benlamine, S., Chaouachi, M., Frasson, C., & Gandon, F. (2017). Emotions and personality traits in argumentation: An empirical evaluation. *Argument & Computation*, *8*(1), 61–87. https://doi.org/10.3233/AAC-170015

Wachsmuth, H., Naderi, N., Habernal, I., Hou, Y., Hirst, G., Gurevych, I., & Stein, B. (2017). Argumentation quality assessment: Theory vs. practice. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, *2*, 250–255. https://doi.org/10.18653/v1/P17-2039

Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., Hirst, G., & Stein, B. (2017). Computational Argumentation Quality Assessment in Natural Language. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, *1*, 76–187. https://aclanthology.org/E17-1017

Wall, J. D., & Warkentin, M. (2019). Perceived argument quality's effect on threat and coping appraisals in fear appeals: An experiment and exploration of realism check heuristics. *Information & Management*, *56*(8), 103157. https://doi.org/10.1016/j.im.2019.03.002

Wang, Q., Yang, X., & Xi, W. (2018). Effects of group arguments on rumor belief and transmission in online communities: An information cascade and group polarization perspective. *Information & Management*, *55*(4), 441–449. https://doi.org/10.1016/j.im.2017.10.004

Watt, S. E., Maio, G. R., Haddock, G., & Johnson, B. T. (2008). Attitude functions in persuasion: Matching, involvement, self-affirmation, and hierarchy. In R. Prislin & W. Crano (Eds.), *Attitudes and attitude change* (pp. 189–211). Psychology Press.

Weber, R., Huskey, R., Mangus, J. M., Westcott-Baker, A., & Turner, B. (2015). Neural predictors of message effectiveness during counterarguing in antidrug campaigns. *Communication Monographs*, *82*(1), 4–30. https://doi.org/10.1080/03637751.2014.971414

Yi, M. Y., Yoon, J. J., Davis, J. M., & Lee, T. (2013). Untangling the antecedents of initial trust in web-based health information: The roles of argument quality, source expertise, and user perceptions of information quality and risk. *Decision Support Systems*, *55*(1), 284–295. https://doi.org/10.1016/j.dss.2013.01.029

Zhao, X., & Cappella, J. N. (2016). Perceived argument strength. In D. K. Kim & J. Dearing (Eds.), *Health communication research measures* (pp. 119–126). Peter Lang.

Zhao, X., Strasser, A., Cappella, J. N., Lerman, C., & Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. *Communication Methods and Measures*, *5*(1), 48–73. https://doi.org/10.1080/19312458.2010.547822

Zillmann, D. (1999). Exemplification theory: Judging the whole by some of its parts. *Media Psychology*, *1*(1), 69–94. https://doi.org/10.1207/s1532785xmep0101_5