International Journal of Communication 13(2019), 1863–1889 1932–8036/20190005

Embedding a Wiki Platform Within a Traditional Survey: A Novel Approach to Assess Perceived Argument Strength in Communication Research

JEFF NIEDERDEPPE[1]
Cornell University, USA

DANIEL A. GUNDERSEN
Dana-Farber Cancer Institute, USA

ANDY S. L. TAN
Dana-Farber Cancer Institute and
Harvard T. H. Chan School of Public Health, USA

EMMA E. McGINTY
COLLEEN L. BARRY
Johns Hopkins Bloomberg School of Public Health, USA

Several prominent theories predict that argument strength plays an important role in message processing and effects. Traditional strategies to measure perceived argument strength have limitations in responsivity to emerging arguments in public discourse. This article examines the utility of a survey-embedded wiki platform (wiki survey) to identify strong and weak arguments in dynamic information environments. Participants completed two wiki surveys, embedded within a larger survey of U.S. adults (N = 1,506), asking them to choose between randomly selected pairs of arguments related to marijuana legalization or to add new arguments to the item pools. The method identified 32 novel, user-generated arguments, over and above an original set of 26 arguments identified by study authors through a review of the literature and a content analysis of news media

Jeff Niederdeppe: jdn56@cornell.edu
Daniel A. Gundersen: DanielA_Gundersen@dfci.harvard.edu
Andy S. L. Tan: andy_tan@dfci.harvard.edu
Emma E. McGinty: bmcginty@jhu.edu
Colleen L. Barry: cbarry@jhu.edu
Date submitted: 2018–10–10

[1] The Survey Research Institute (SRI) at Cornell University conducted all survey programming and data collection. We are grateful to Nicholas Brennan, Heather Bianchi, and Kathryn Varga for developing a codebook and coding all open-ended survey responses, and to Heather Bianchi and Mia Jovanova for assistance in preparing and formatting data tables. The authors have no conflicts of interest to report.

coverage on the topic. The wiki survey also produced variation in perceived strength of arguments among relevant demographic and social groups.

Keywords: elaboration likelihood model, competitive framing theory, message effects, persuasion, health communication, public opinion

Researchers have devoted considerable attention to methods that seek to identify strong arguments in favor of, or in opposition to, issues of public interest and concern. Several theories, including the elaboration likelihood model (ELM; Petty & Cacioppo, 1986) and competitive framing theory (CFT; Chong & Druckman, 2007a), predict that perceived argument strength (PAS) plays an important role in the processing and effects of strategic messages. There has also been substantial interest in using self-reported measures of perceived message effectiveness (PME), a related concept, in understanding the persuasive effects of strategic messages (e.g., Dillard & Ye, 2008; O'Keefe, 2018; Yzer, LoRusso, & Nagler, 2015). Traditional strategies to measure these concepts have considerable trade-offs in terms of researcher and respondent burden and their ability to respond to a rapidly changing information environment in a timely manner.

This article examines the utility of a survey-embedded wiki platform (a wiki survey) to identify strong arguments in support of and opposition to social policies. We explore the pros and cons of wiki surveys to gauge perceived strength of arguments for and against legalization of marijuana for recreational use, a policy that has generated considerable public debate and momentum in recent years. We compare results from wiki surveys with those from traditional survey methods to gauge PAS to assess benefits and trade-offs involved with wiki surveys.

Theoretical Background

Two major theories, the ELM and CFT, consider the interaction between audience motivation and message content as key factors in shaping the likelihood, direction, and magnitude of persuasive effects. Both theories also predict a central role for PAS in shaping how motivated audiences process and respond to messages.

The ELM emphasizes that argument strength is likely to matter under conditions when audiences are motivated and able to engage in effortful thought and mental deliberation (Petty & Cacioppo, 1986). Under such conditions, audiences will engage in deep processing of a message and carefully scrutinize an argument's quality and evidence. Strong arguments (defined as those that generate positive thoughts among members of the audience) are likely to produce enduring changes in attitudes that persist over time, whereas weak arguments are unlikely to persuade and may even catalyze unintended, counterattitudinal responses. Although the ELM predicts that argument strength is unlikely to matter in cases where audiences lack the motivation or ability to process a message, argument strength is theorized to be the most important factor in predicting whether a message succeeds or backfires among audiences who are willing and able to devote time and effort to process the message. A recent meta-analysis, analyzing 134 studies, confirmed this basic proposition (Carpenter, 2015).

CFT introduces a similar concept, frame strength, in predicting the relative success or failure of competing messages in shaping attitudes (Chong & Druckman, 2007a). CFT describes frame strength as "increasing with the persuasiveness of a given frame" (Chong & Druckman, 2007a, p. 103), and CFT studies measure the construct by asking audience members to identify whether a series of arguments are "effective" at supporting a policy proposal about a social issue (Chong & Druckman, 2007b, p. 641). Thus, frame strength appears to describe a concept nearly identical to that of argument strength. CFT argues that frame strength will matter most in situations where audiences are (a) exposed to competing messages—some in support of and others in opposition to a policy or position—and (b) motivated by prior interest or involvement in the issue to engage in effortful scrutiny of arguments and their evidence (Chong & Druckman, 2007a). Overall, CFT predicts that strong frames will outperform weak frames in competitive contexts among all audiences. Furthermore, although weak frames can persuade low-motivated audiences when offered in isolation, they run the risk of backfiring among highly motivated audiences, particularly in the context of concurrent exposure to strong counterframes.

The ELM and CFT make clear that theoreticians, applied researchers, and strategic communicators would benefit from the ability to distinguish, a priori, arguments perceived to be strong or weak among relevant audiences. However, although there are many options for measuring PAS, each has notable limitations in terms of respondent burden, researcher burden, and responsiveness to the emergence of new arguments in public discourse, as noted below.

Traditional Strategies to Gauge PAS

Most studies that attempt to characterize argument strength, including those informed by the ELM and CFT, characterize argument strength in terms of audience perceptions rather than intrinsic qualities of arguments themselves. Although we acknowledge a historically rich and vast literature in rhetoric and argumentation that explores intrinsic qualities of strong arguments (Else, 1957; Walton, Reed, & Macagno, 2008), we focus on strategies designed to assess PAS to inform persuasion research.

Thought-Listing Methods

The ELM's authors proposed a thought-listing procedure to gauge PAS in research testing the theory (Petty & Cacioppo, 1986). Thought listing presents respondents with a series of arguments and asks them to speak or write the thoughts that came to mind as they read each argument. Researchers then train independent judges, or ask respondents themselves, to code each thought for its valence (Tarico, Van Velzen, & Altmaier, 1986). Researchers aggregate responses for each argument and assess the volume or ratio of positive and negative thoughts. Arguments that generate predominantly positive thoughts are considered strong, whereas arguments that generate largely negative thoughts are considered weak.

Thought listing has many strengths as a way to characterize PAS. Studies find stronger message effects among highly motivated audiences when researchers gauge argument strength using the thought-listing procedure versus other methods (Carpenter, 2015). Open-ended responses also provide detailed information about audience responses that researchers can code along dimensions like origin (internal or external to the message) or target (the claim, the evidence, the source, the channel; Tarico et al., 1986). The

method has major limitations, however, as noted elsewhere (Zhao, Strasser, Cappella, Lerman, & Fishbein, 2011). Most notable are respondent burden (writing or typing their responses verbatim, a time-intensive task that produces large variance in response length) and researcher burden (training a team of coders to reliably assess the valence of these responses, an iterative process time is time and labor intensive). This may explain why most studies testing the ELM use closed-ended scales rather than thought listing to gauge argument strength (Carpenter, 2015). Although computer-automated machine learning and/or sentiment analysis could reduce researcher burden associated with coding thought-listing data in the future, it is yet unclear that they offer the same degree of predictive validity (Baek, Cappella, & Bindman, 2011).

Closed-Ended Scales of PAS or PME

The most common technique used to gauge argument strength uses closed-ended, scaled items. This method presents respondents with a series of arguments and asks them a series of Likert or semantic differential scaled items about the argument. A closely related literature has developed around the concept of PME, which uses a similar procedure, but presents respondents with messages that may have multiple arguments and/or persuasive appeals (e.g., a 30-second antismoking advertisement; see Yzer et al., 2015).

There is substantial variation in the number and specific content of these items used in practice, ranging from a single item (e.g., Chong & Druckman, 2007a; Gollust, Niederdeppe, & Barry, 2013) to 9-item scales (Zhao et al., 2011). Studies following the theoretical traditions of the ELM and CFT often ask respondents to assess the effectiveness of an argument in making the case for or against a particular policy proposal, scaling these items from *definitely not effective to definitely effective* (e.g., Chong & Druckman, 2007b; Petty & Wegener, 1998). Zhao et al. (2011) validated a scale that features various different dimensions of PAS that include perceptions of the valence of thoughts generated, believability, importance, helpfulness, and whether or not respondents agree with the argument in the first place. Other researchers have used items that vary the dimension of evaluation (e.g., cognitive: persuasive, logical or rational; emotional: feel sad, excited or scared), the referent (e.g., effective among myself, close friends, people like me, people in general), and the raters themselves (e.g., target audience, experts; Dillard & Ye, 2008; O'Keefe, 2018; Yzer et al., 2015).

Closed-ended scales also have notable strengths as a strategy to gauge argument strength. Zhao et al. (2011) validated their scale in several settings, and the scale correlates strongly with thought-listing results. Analyzing closed-ended items also reduces researcher burden. There are several notable limitations to this approach, however. Meta-analytic data suggest that strong arguments identified via closed-ended scales are less predictive of the ELM's central prediction (that strong arguments matter more for highly motivated/able audiences) than strong arguments identified via thought listing (Carpenter, 2015). O'Keefe (2018) raised similar questions about whether closed-ended assessments of PAS are reliably predictive of attitude change. Some of these challenges may stem from the fact that researchers use vastly different dimensions of evaluation, referents, and raters in these assessments (Dillard & Ye, 2008; Yzer et al., 2015). Multi-item measures can also be burdensome to participants if studies seek to assess the strength of a large number of arguments.

Limitations of Both Thought-Listing and Closed-Ended Scales

Both methods require the researchers to possess, a priori, a comprehensive list of all candidate arguments to test them. However, CFT makes clear that public debates around social issues and policy solutions are dynamic, such that competing actors continuously develop and disseminate new arguments to counter arguments offered by the opposition or in response to new data or evidence (Chong & Druckman, 2013). Novel arguments may also develop in an effort to persuade particular issue publics or demographic groups. Thus, both traditional thought-listing and closed-ended scale-based strategies to characterize the strength of a comprehensive set of possible arguments run the risk of missing key considerations in the absence of a mechanism to collect immediate audience feedback and incorporate that feedback into the testing procedure. Some researchers try to address this problem by identifying a large set of available arguments by analyzing the public information environment (e.g., news coverage) of social issues before testing argument strength (e.g., Gollust et al., 2013; McGinty et al., 2016). These content analyses, however, are sufficiently labor intensive and time-consuming, such that new arguments may emerge while these data are being collected and analyzed. It should be noted that in-depth, qualitative interviews and focus groups are designed to provide this kind of immediate feedback and are often used in real-world campaign development (Atkin & Freimuth, 2001; Fishbein & Azjen, 2009). These methods, however, are difficult to scale and may not provide clear distinctions in argument strength across demographic subgroups.

Wiki Surveys: A Potential Strategy to Address These Limitations

Salganik and Levy (2015) developed a novel data collection approach, pairwise wiki surveys (hereafter "wiki surveys"), in response to these limitations. Wiki surveys combine the benefits of collecting user-generated content and quantitative surveys among large samples, which, if drawn from a known-probability method, can produce generalizable rankings. A wiki survey involves randomly pairing two items from a pool of items and asking a randomly selected participant to choose their preferred item, or to contribute a new item. User-generated items are then moderated by an administrator and added to the pool of items. Using this approach, data collected from multiple pairwise comparisons from many participants can be analyzed to determine the relative rankings of each item across the entire set of items.

An example of its application can be seen in Salganik and Levy's (2015) work identifying New York City residents' suggestions for the city's sustainability plans. An initial set of items was generated by the research team and the New York City Mayor's Office of Long-Term Planning and Sustainability. They found that eight of the 10 most highly rated suggestions were user generated. The authors reported similar findings for the Organization for Economic Cooperation and Development's "Raise Your Hand" initiative to obtain input from stakeholders within the global education community on the most important actions for improving education in the 21st century. They concluded that the wiki survey offered advantages that would not have been possible with closed-ended surveys, focus groups, or qualitative interviews.

Neither of the studies described above used a wiki survey to identify strong versus weak arguments, nor did they embed the wiki platform within a traditional survey to permit ranking comparisons by individual characteristics. However, the logic of the approach seems well suited to address the limitations of thought-listing and closed-ended scale procedures for measuring PAS. Thus, we sought to explore the utility of a

wiki survey to evaluate the relative strength of existing arguments, as well as to identify new arguments, in support of or opposition to a rapidly evolving policy issue in the United States: recreational marijuana legalization.

The Context: Recreational Marijuana Legalization

Both public opinion and the regulatory status of recreational marijuana have changed dramatically in recent years. In January 2014, Colorado became the first state in the U.S. to make it legal to sell recreational marijuana for adult use. Three additional states (Washington, Oregon, and Alaska) legalized recreational marijuana between 2014 and 2015, and the District of Columbia made it legal for adults to possess and grow marijuana beginning in February 2015. Four more states (California, Maine, Massachusetts, and Nevada) voted to legalize recreational marijuana in November 6, 2016, ballot initiatives. These shifts have occurred against the backdrop of rapidly evolving public views and policy changes, including passage of medical marijuana laws in more than half of U.S. states over the previous decade. A majority of American adults now believe that marijuana should be made legal, a remarkable shift from the 1980s, when public support was less than 20% (Geiger, 2016; McGinty, Niederdeppe, Heley, & Barry, 2017).

Despite these changes, there remains considerable public debate about the benefits and costs of legalization. Topics of debate include impacts on public health and safety, law enforcement, the economy, and the criminal justice system. Content analysis indicates that journalists have offered a balanced portrayal of the pros (53%) and cons (47%) of legalization (McGinty et al., 2016). Prominent arguments emphasize the benefits of reduced burden on the criminal justice system and the potential for increased tax revenue while also considering possible adverse public health consequences to youth (addiction) and adults (impaired driving). A recent public opinion survey found that American adults are optimistic about potential benefits to the economy, but concerned about health consequences and worried about conflicts between state and federal laws about the legality of recreational marijuana use (McGinty et al., 2017).

Recreational marijuana appears to represent a competitive and dynamic social issue that is ripe for rapid evolution of arguments in public discourse, a topic for which a wiki survey would seem particularly relevant. We also chose the topic based on the availability of data about (a) the types of arguments that appeared in public discourse on the topic in recent years (McGinty et al., 2016) and (b) the perceived strength of those arguments using traditional, closed-ended scales on a public opinion survey (McGinty et al., 2017). This provided the research team with a large set of candidate arguments to form the initial pool of options, and in doing so, offered a conservative test of the degree to which the wiki survey approach can identify novel arguments.

Research Questions

We offer four research questions to explore whether a survey-embedded wiki survey offers unique potential to gauge PAS for competitive and dynamic public policy issues. In light of available, closed-ended scale data on PAS of a wide variety of arguments for and against recreational marijuana legalization that have appeared in the public discourse in recent years (McGinty et al., 2017), we first ask how many novel, user-generated arguments will be generated by the wiki survey approach.

RQ1: How many novel, user-generated arguments did participants generate?

A second set of questions concerns the degree to which wiki surveys offer information that is comparable to traditional, closed-ended scale approaches, and whether the wiki survey approach can efficiently differentiate higher performing from lower performing items. One potential benefit of the wiki survey approach is that items can receive many ratings because respondents are asked to rate pairs of items as long as she or he chooses to engage in the task; at the same time, respondents could also choose to respond to fewer comparisons. McGinty et al. (2017) considered 26 different arguments about marijuana legalization (13 pro and 13 anti), which reflects many possible pairwise comparisons. We thus ask how strongly the wiki survey rankings correlate with the rankings observed in prior research using traditional survey methods (McGinty et al., 2017).

RQ2: How strongly do wiki survey rankings correlate with scale measures?

A final pair of questions concerns the degree to which wiki surveys differentiate perceptions of argument strength among relevant social groups. The ability to differentiate argument strength among relevant subpopulations would be of value both to understand the underlying nature of public opinion about specific social issues and to identify promising message strategies that could be targeted to relevant subpopulations.

Relevant social groups are likely to differ depending on the issue at hand, but state legalization status and a priori policy support are likely important for recreational marijuana legalization. Although McGinty et al. (2017) did not observe striking differences in the relative strength of arguments by state legalization status, this could reflect the fact that traditional, closed-ended survey methods gauge PAS in absolute rather than relative terms, and thus may confound policy experience or existing policy support with argument strength. This raises the possibility that the wiki survey approach, which assesses argument strength in only relative terms, may offer greater nuance in identifying arguments that are uniquely compelling among relevant social groups. Hornik and Woolf (1999) further argue that the ability to distinguish between strong and weak arguments in predicting differences in the targeted outcome (support for legalization) offers valuable information to inform the development of strategic messages aimed at increasing (or decreasing) support. We thus ask whether the wiki survey identifies different strong and weak arguments by state legalization status (RQ3), and support of, versus opposition to, the policy (RQ4).

RQ3: Does the wiki survey differentiate strong and weak arguments by state legalization status?

RQ4: Does the wiki survey differentiate strong and weak arguments by policy support?

Method

Survey Procedure

We contracted with the Survey Research Institute at Cornell University to recruit a national sample of U.S. adults ages 21 and older from an online survey research panel. A sample of panel members living

in the U.S. received an e-mail invitation to participate in an online study about a "controversial social issue" in exchange for small incentive ($0.40 to $0.60, depending on the number of other surveys they had recently completed as part of the panel). Participants who provided informed consent then answered a series of questions about their state of residence (to assess state legalization status), demographics, political ideology, and whether they supported or opposed recreational marijuana legalization. Data collection occurred between October 4 and November 2, 2016.

Embedded in the survey were two wiki surveys—contrasting prolegalization arguments and antilegalization arguments—administered in random order. In comparing arguments, respondents could choose a tie ("I can't decide") or add their own arguments by clicking the "add your own idea" button. Respondents could complete as many preference votes they liked. After two minutes, the "next" button appeared in the lower right corner of the survey, and respondents had the option to proceed to the other wiki survey (antilegalization arguments if they were randomly assigned to prolegalization arguments first, and vice versa). A total of 1,506 answered all of the demographic questions and participated in at least one of the two wiki surveys; the overall response rate was 10% (American Association for Public Opinion Research (AAPOR) response-rate calculation #4). Respondents voted 92,695 times on pairs of pro-legalization arguments (*Mdn* = 14, *M* = 55.5 votes per respondent; range: 1–1,919) and 65,139 times on pairs of antilegalization arguments (*Mdn* = 10, *M* = 42.8 votes per respondent; range: 1–1,390).

A team of trained undergraduate research assistants monitored all user-generated submissions daily to filter out irrelevant or offensive commentary, identify novel arguments, and rewrite them to be comparable in structure (typically starting with "legalization would . . .") and evidence (each argument could have no more than one fact or statistic). New arguments that met these inclusion criteria were added to the pool of candidate items for all future survey iterations. We split double-barreled but otherwise eligible arguments into separate items and added them to the pool. The wiki survey platform weights new items so that respondents see user-generated arguments more frequently than existing arguments. Thus, item sample sizes converge over time.

Sample Characteristics

We oversampled respondents using quotas similar to those employed by McGinty et al. (2017) to ensure demographic comparability among the studies. Our target sample size was 1,500, roughly a third of which we oversampled from states or jurisdictions that had already passed recreational legalization (AK, CO, DC, OR, and WA). We thus began the survey by asking respondents which state they resided in. We also sought equal proportions of Democrats, Republicans, and Independents, and demographics that approached U.S. Census demographics. We oversampled as necessary and feasible by political ideology, age, gender, education, race, and ethnicity. To do so, we gauged demographic characteristics using the same wording of these measures employed by McGinty et al. (2017) and measured political ideology by asking, "Generally speaking, do you think of yourself as a Republican, Democrat, Independent, another party, or no preference?" We restricted data collection to 50 respondents per day to ensure that new arguments would have enough responses from which to gauge their strength.

The sample was majority male (59%), largely White (89%), non-Hispanic (93%), well-distributed by age, highly educated (86% with at least some college education), and relatively evenly stratified by political party affiliation (28% Republican, 27% Independent, 38% Democrat, 7% Other). These proportions differ somewhat from those obtained by McGinty et al. (2017), most notably in terms of sex (McGinty et al.'s sample was 55% female) and education (McGinty et al.'s sample had only 64% with some college).

Ranking of PAS

We began the wiki survey portion of the study with the following instructions: "Many people have offered arguments for why they think legalizing marijuana for recreational use by adults is a [good/bad] idea. In this section of the survey, we would like for you to choose between two arguments for why some people think legalizing marijuana for recreational use by adults is a [good/bad] idea. For each pair, please read both of the statements and choose which statement you think is a better reason to [support/oppose] legalization of marijuana for recreational use by adults." After clicking the "next" button, respondents proceeded to a page that embedded the wiki survey within the larger survey platform and responded to as many argument pairs as desired.

We fit a Bradley–Terry model (using the BradleyTerry2 package in R; R Core Team, 2017; Turner & Firth, 2012) to generate relative ranking of statements, using the median performing statement as the reference category. In short, Bradley–Terry models are a special case of generalized linear models that estimate the log odds that one item will outperform another in a pairwise comparison. Setting one statement (in our case, the median) as the reference item (i.e., log odds = 0), the coefficients can be interpreted as the expected log odds of winning for a statement against the median performing statement. The relative magnitude of the coefficients can then be used to rank the items from best performing (largest positive value) to worst performing (smallest negative value). Pairs of items could repeat for a respondent, so we used cluster sandwich standard errors to adjust for correlation of votes within respondents using the sandwich package in R (Zeileis, 2004, 2006). We categorized each argument as original (included by McGinty et al., 2017) or user generated to permit comparisons.

Support for Recreational Marijuana Legalization

We gauged support for recreational marijuana legalization using wording identical to that employed in McGinty et al. (2017) to ensure comparability. For those living in CO, WA, OR, AK, or DC, we asked, on a 7-point Likert scale, from *strongly oppose* to *strongly support*, "Marijuana for recreational use by adults was legalized by your state in [YEAR]. Do you support or oppose this law?" For those living in all other states, we asked (on the same scale), "Do you support or oppose legalizing marijuana for recreational use by adults in the state where you currently live?" A majority of respondents in marijuana-legalized states (64%) and all other states (59%) offered support (values above the midpoint, including *somewhat support*, *support*, or *strongly support*) for legalization in their state.

Results

RQ1: How many novel, user-generated arguments did participants generate?

Respondents offered 285 responses by clicking on the "add your own idea" button on the prolegalization wiki survey page, and 356 responses to the same prompt on the antilegalization arguments page. We trained three undergraduate research assistants to reliably code these responses (Krippendorf's alpha > .80 for both pro- and antilegalization ideas) and identify discrete and novel arguments that did not directly overlap with ideas already included in the pool. This process identified 11 novel prolegalization arguments and 21 new antilegalization arguments. We reworded these items for consistency and included them in the available pool of arguments presented to remaining participants. Although the rest of the open-ended comments took many forms, the most common substantive responses involved respondents stating their opinion without engaging with an argument to support it (e.g., "I do not condone drug use"; "I always support marijuana legalization"; *n* = 90, 14% of responses), noting that they reject both arguments (*n* = 77, 12% of responses), offering an argument that duplicated one in the pool already (*n* = 73, 11% of responses), and commenting on negative experiences taking the wiki survey (e.g., "biased survey question," "this survey keeps going in an endless loop"; *n* = 30, 5% of responses).

RQ2: How strongly do Wiki survey rankings correlate with scale measures?

Table 1 and Table 2 list all pro- and antilegalization arguments in the wiki survey item pools, including both original and user-generated items, ranked according to the Bradley–Terry models. We also show levels of agreement from the items also measured by McGinty et al. (2017) to facilitate comparisons in the rankings between the two methods and studies.

Table 1. Prolegalization Arguments From the Wiki Survey and McGinty et al. (2017).

Prolegalization Arguments		% Agree [95% CI],[2] McGinty et al. (2017)	Log Odds vs. Median Rank [95% CI][3], Wiki Survey	Original or User Generated
Legalization would . . .				
Lower law enforcement costs	Save $8 billion dollars across the U.S. in law enforcement costs currently spent on enforcing marijuana prohibition	61.3 [58.3, 64.4]	0.85 [0.67, 1.03]	Original
Reduce prison overcrowding	Reduce prison overcrowding by ending the imprisonment of nonviolent marijuana users	62.8 [59.7, 65.7]	0.74 [0.58, 0.91]	Original
Provide medical treatment	Be helpful for people with medical conditions		0.71 [0.47, 0.94]	User

[2] Data collection occurred between April 15 and April 28, 2016.

[3] Data collection occurred between October 4 and November 2, 2016.

Reallocate law enforcement resources	Allow law enforcement resources to be more appropriately allocated		0.60 [0.43, 0.77]	User
Increase tax revenue	Increase tax revenue by as much as $6 billion per year across the U.S.	63.9 [60.9, 66.9]	0.50 [0.31, 0.70]	Original
Enable medical marijuana research	Clear hurdles for medical marijuana research		0.49 [0.28, 0.69]	User
Reduce criminal justice spending	Reduce the amount of money spent on the criminal justice system		0.47 [0.31, 0.63]	User
Counter drug cartels	Reduce the power and profits of the drug cartels that control the majority of the marijuana market	57.7 [54.5, 60.8]	0.31 [0.12, 0.49]	Original
Create a profitable new industry	Create a new industry worth $14 billion in sales each year across the U.S.	63.4 [60.3, 66.4]	0.25 [0.04, 0.46]	Original
Increase job production	Create 500,000 new jobs across the U.S.	45.7 [42.7, 48.9]	0.15 [−0.06, 0.37]	Original
Reduce prescription medication abuse	Reduce the abuse of prescription medication		0.07 [−0.73, 0.86]	User
Reduce violent crime	Reduce violent crime by closing down illegal marijuana markets	42.3 [39.3, 45.5]	0.05 [−0.12, 0.23]	Original
Provide consistent law enforcement	Allow for consistent enforcement of the law across all states		Median ranked item (0.00)	User
Reduce youth access	Reduce youth access to marijuana by replacing illegal dealers with legal stores that would not sell marijuana to minors	44.4 [41.3, 47.6]	−0.04 [−0.24, 0.16]	Original

Protect users from dangerous illegal markets	Protect marijuana users from dangerous illegal drug markets by moving marijuana sales to stores like any other product	52.3 [49.2, 55.4]	−0.07 [−0.25, 0.12]	Original
Improve product safety	Protect marijuana users from consuming unsafe or tainted marijuana products	55.9 [52.7, 58.9]	−0.22 [−0.42, −0.05]	Original
Decrease street gang income	Eliminate a means of income for street gangs		−0.23 [−0.42, −0.05]	User
Reduce overdose deaths	Reduce overdose deaths from prescription painkillers and heroin	31.9 [29.0, 34.9]	−0.42 [−0.42, −0.03]	Original
Reduce racial inequalities	Reduce racial inequalities in nonviolent drug arrests and convictions	39.2 [36.2, 42.3]	−0.46 [−0.62, −0.30]	Original
Enforce age limits	Allow states to enforce age limits, as with cigarettes		−0.49 [−0.66, −0.33]	User
Allow the right of recreational use	Allow for the use of recreational marijuana, a basic personal right		−0.74 [−1.01, −0.48]	User
Provide a safer alternative to alcohol	Allow individuals to enjoy using marijuana instead of alcohol, a harmful substance		−0.87 [−1.11, −0.64]	User
Reward productive workers	Allow productive workers to capitalize on wages forfeited by marijuana users who come to work under the influence		−1.58 [−1.81, −1.34]	User
Increase tourism revenue	Increase tourism revenue	35.3 [32.4, 38.4]	−1.59 [−1.59, −1.33]	Original

Table 2. Antilegalization Arguments From the Wiki Survey and McGinty et al. (2017).

Antilegalization Arguments		% Agree [95% CI],[4] McGinty et al. (2017)	Log Odds vs. Median Rank [95% CI],[5] Wiki Survey	Original or User Generated
Legalization would . . .				
Increase youth access	Increase access to marijuana by youth, putting their long-term health at risk	49.6 [46.5, 52.8]	0.55 [0.35, 0.75]	Original
Increase motor vehicle crashes	Increase rates of car crashes due to marijuana-impaired driving	51.8 [48.6, 54.9]	0.51 [0.30, 0.73]	Original
Increase accidental overdose	Lead to accidental overdose among children who eat marijuana edibles, like cookies or candy	47.0 [43.9, 50.1]	0.51 [0.31, 0.71]	Original
Increase use of other drugs	Increase use of other illegal drugs. Youth will try marijuana and then move on to other drugs like heroin or cocaine	45.7 [42.7, 48.9]	0.46 [0.23, 0.69]	Original
Reduce workplace productivity	Lead to more people coming to work under the influence of marijuana, decreasing productivity in the workplace	51.5 [48.3, 54.6]	0.45 [0.23, 0.66]	Original
Increase health-care spending	Increase U.S. health-care spending by $11 billion dollars due to increases in hospital visits and addiction treatment	39.8 [36.8, 42.9]	0.45 [0.23, 0.66]	Original
Increase crime	Lead to increases in crime. Like many alcohol outlets, marijuana business will become magnets for criminal activity	42.6 [39.5, 45.7]	0.29 [0.09, 0.48]	Original

[4] Data collection occurred between April 15 and April 28, 2016.

[5] Data collection occurred between October 4 and November 2, 2016.

Expand marijuana industry	Create a powerful and harmful industry that will aggressively market its dangerous product to young people	49.5 [46.4, 52.7]	0.28 [0.07, 0.50]	Original
Harm physical and mental health	Be detrimental to the physical and mental health of the U.S.		0.20 [−0.06, 0.47]	User
Conflict with federal law	[Different stem language] Legalization by states when the drug is illegal under federal law creates problems for state lawmakers, businesses, and citizens	63.0 [59.9, 66.0]	0.18 [−0.02, 0.37]	Original
Jeopardizes public safety	Be bad for public safety		0.17 [−0.09, 0.43]	User
Cause unpredictable behavior	Allow easier access to a drug that causes unpredictable behavior among its users		0.17 [−0.16, 0.50]	User
Increase rates of dependence	Cause more individuals to become dependent on marijuana.		0.11 [−0.16, 0.38]	User
Reduce self-control and responsibility	Encourage a lack of self-control and personal responsibility		0.10 [−0.16, 0.35]	User
Gateway drug	[Different stem language] Marijuana is a gateway drug		0.09 [−0.18, 0.35]	User
Harm at any age	[Different stem language] Legalization of a harmful drug is not something that should be legal for anyone, at any age		0.03 [−0.27, 0.32]	User
Increase cancers	Lead to increased rates of lung and throat cancers		Median ranked item (0.00)	User

Increase neighborhood inequalities	Lead to concentration of marijuana retailers in low income and minority neighborhoods	42.2 [39.1, 45.3]	−0.05 [−0.25, 0.16]	Original
Perpetuate addiction	Allow people to feel their addictions		−0.07 [−0.32, 0.18]	User
Lack of research	[Different stem language] There is a lack of research on the effects of marijuana use		−0.14 [−0.41, 0.13]	User
Increase in birth defects	Lead to an increase in birth defects among newborns, offsetting any potential benefits		−0.16 [−0.38, 0.05]	User
Fail to eliminate the black market	Not eliminate the black market	57.2 [54.1, 60.3]	−0.16 [−0.40, 0.08]	Original
Allow tobacco and pharma company involvement	Allow tobacco and pharmaceutical companies to become involved in the marijuana industry, driving up costs		−0.23 [−0.55, 0.08]	User
Threaten moral values	Threaten the moral values that make America strong	40.9 [37.8, 44.0]	−0.35 [−0.65, −0.05]	Original
Threaten the environment	Be bad for the environment. The marijuana industry will pollute the earth with pesticides and use up already scarce water resources	30.8 [28.0, 33.8]	−0.57 [−0.79, −0.36]	Original
Only eliminate black market	Only eliminate the black market for marijuana		−0.72 [−0.95, −0.49]	User
Modern prohibition	[Different stem language] Laws banning marijuana are just a modern form of prohibition, an idea that failed before		−0.79 [−1.03, −0.55]	User

Increase cost	Cause marijuana to become more expensive if sales were controlled by the government	−0.84 [−1.06, −0.62]	User
Increase intrusive ads	Lead to an increase in intrusive advertisements	−0.89 [−1.12, −0.66]	User
Increase alcohol advertisement	Lead to an increased advertisement of alcohol products as adults reduce their alcohol consumption in favor of marijuana	−0.90 [−1.10, −0.69]	User
Downfall of civilization	Greatly contribute to the downfall of our civilization, allowing strong members of society to dominate the weak	−0.94 [−1.25, −0.64]	User
Age-limit reinforcement	Allow states to enforce age limits, as with cigarettes	−0.96 [−1.19, −0.73]	User
Decrease prison/law enforcement profits	Cut into prison and law enforcement profits	−1.28 [−1.51, −1.04]	User
Slows population growth	Serve as a form of natural birth control, slowing population growth	−1.65 [−1.97, −1.33]	User

We make two primary observations about these data. First, four of the eight strongest prolegalization arguments were user generated, but none of the user-generated antilegalization arguments were among the eight strongest. Still, the wiki survey revealed a wider range and diversity of antilegalization arguments than McGinty et al. (2016) uncovered in a previous content analysis of news media coverage of the issue.

Second, ratings of PAS for arguments that were included in both studies were similar for prolegalization arguments, but quite different for oppositional claims. We correlated the percentage agreement scores from McGinty et al. (2017) with the log-odds coefficients from the current study for the 13 pro- and 13 antilegalization arguments that were included in both studies. This served as a rough proxy for the degree of correspondence between the two ratings methods. The Pearson correlation coefficient for prolegalization arguments was $r = .77$, indicating that the wiki survey item scores accounted for nearly 60% of the variance in percentage agreement scores for prolegalization arguments. Descriptively, three of the top four rated prolegalization arguments from McGinty et al. (2017) were also among the top eight rankings

that emerged in the wiki survey, and the three lowest rated prolegalization arguments from McGinty et al. were also among those with the lowest relative rankings. The biggest discrepancies emerged for arguments related to legalization's effects on "improve product safety" (perceived as a stronger argument in McGinty et al. than in the wiki survey) and "increase job production" (perceived as a stronger argument in the wiki survey than in McGinty et al.).

Ratings for the strongest antilegalization arguments differed substantially between the two methods. The correlation for antilegalization arguments was $r = .39$, indicating that the wiki survey item scores accounted for only 15% of the variance in percentage agreement scores for antilegalization arguments. Descriptively, the strongest antilegalization argument in McGinty et al. (2017), "conflict with federal law," did not emerge among the top eight oppositional arguments in the wiki survey. McGinty et al.'s second strongest antilegalization argument, pertaining to "failure to eliminate black markets" for the drug, was not even in the top half of wiki survey-rated oppositional arguments and performed below the median-ranked argument. Nevertheless, the other five arguments that were in the top half of McGinty et al.'s rankings also appeared in the top eight-ranked arguments in terms of the wiki survey log-odds coefficients.

RQ3: Does the wiki survey differentiate strong and weak arguments by state legalization status?

Figure 1 shows the log odds of winning against the median argument among prolegalization arguments by state legalization status at the time of the survey. Figure 2 shows parallel information for antilegalization arguments. Although differences in PAS were not dramatic, the data suggest greater differentiation between prolegalization arguments among residents of states that have legalized marijuana than states who have not. This can be seen for a few weak arguments (e.g., "improve product safety," "enforce age limits," "reward productive workers") where log odds are more negative relative to the median. There were no clear patterns of difference by state legalization status in the perceived strength of antilegalization arguments.

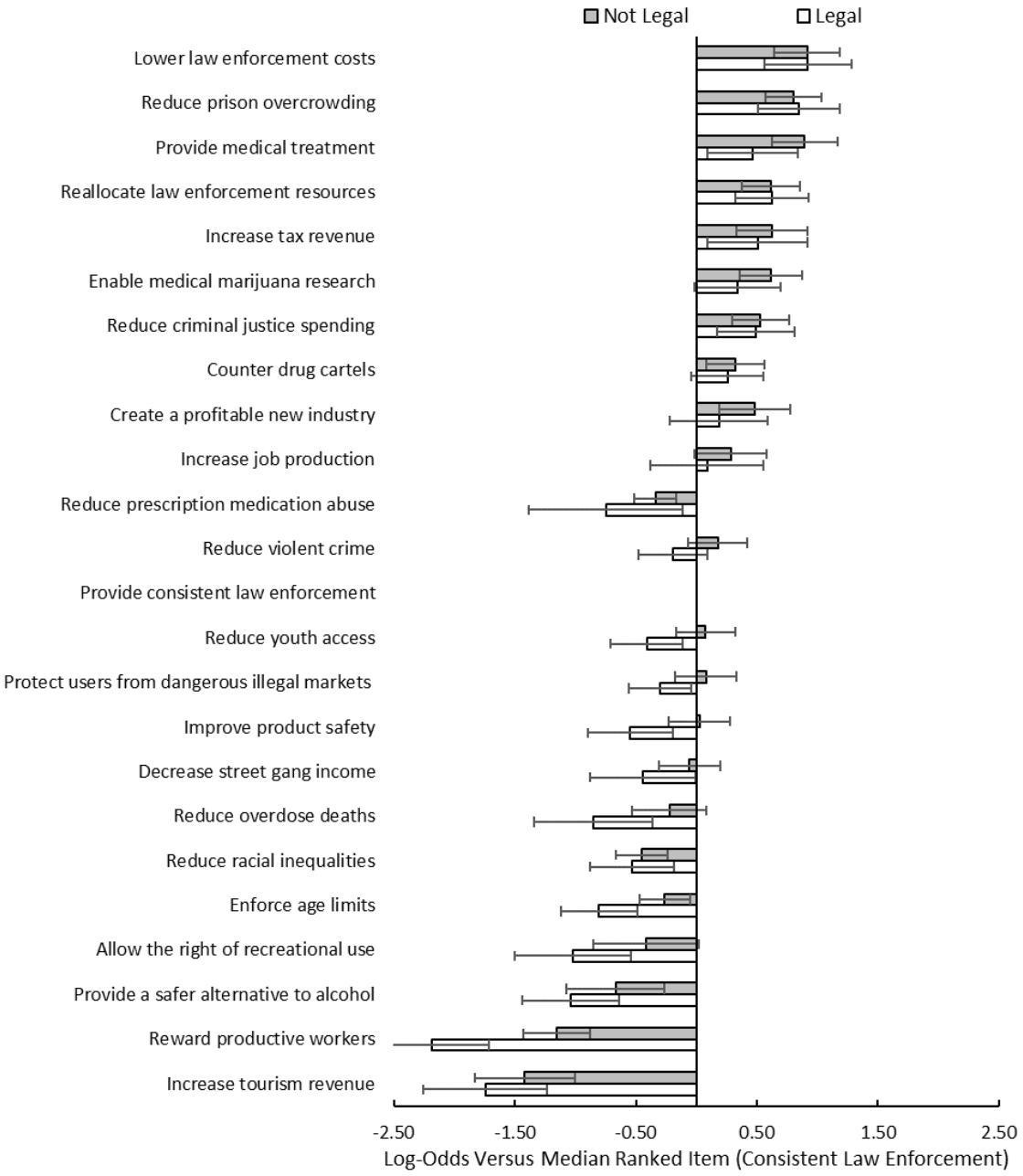

Figure 1. Relative strength of prolegalization arguments by state legalization status.

Note: Error bars depict the 95% CI for the estimated log odds of winning for a statement against the median performing statement

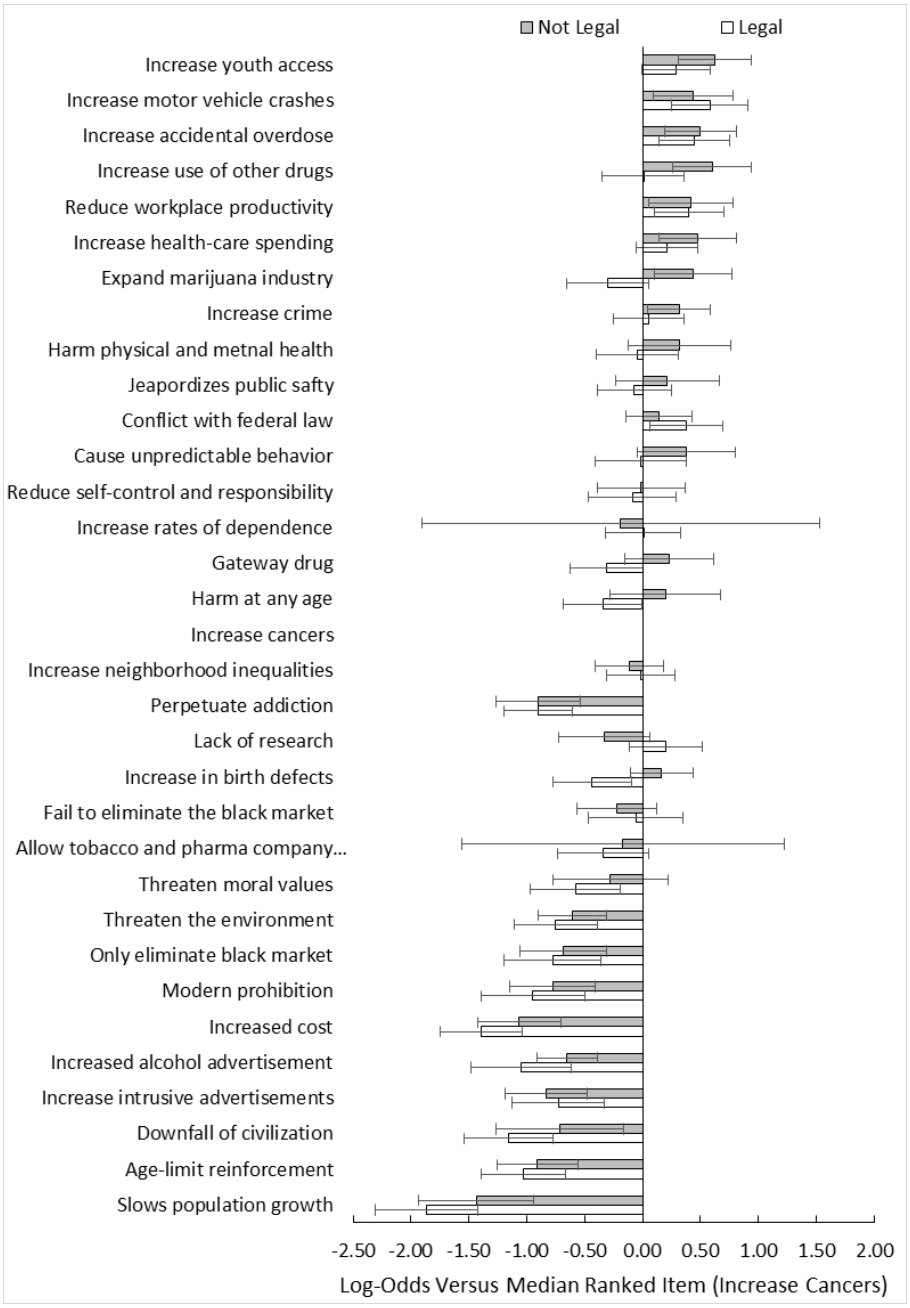

Figure 2. Relative strength of antilegalization arguments by state legalization status.

Note: Error bars depict the 95% CI for the estimated log odds of winning for a statement against the median performing statement.

RQ4: Does the wiki survey differentiate strong and weak arguments by policy support?

Figure 3 depicts the log odds of winning against the median argument for prolegalization arguments by respondent support or opposition to legalization.

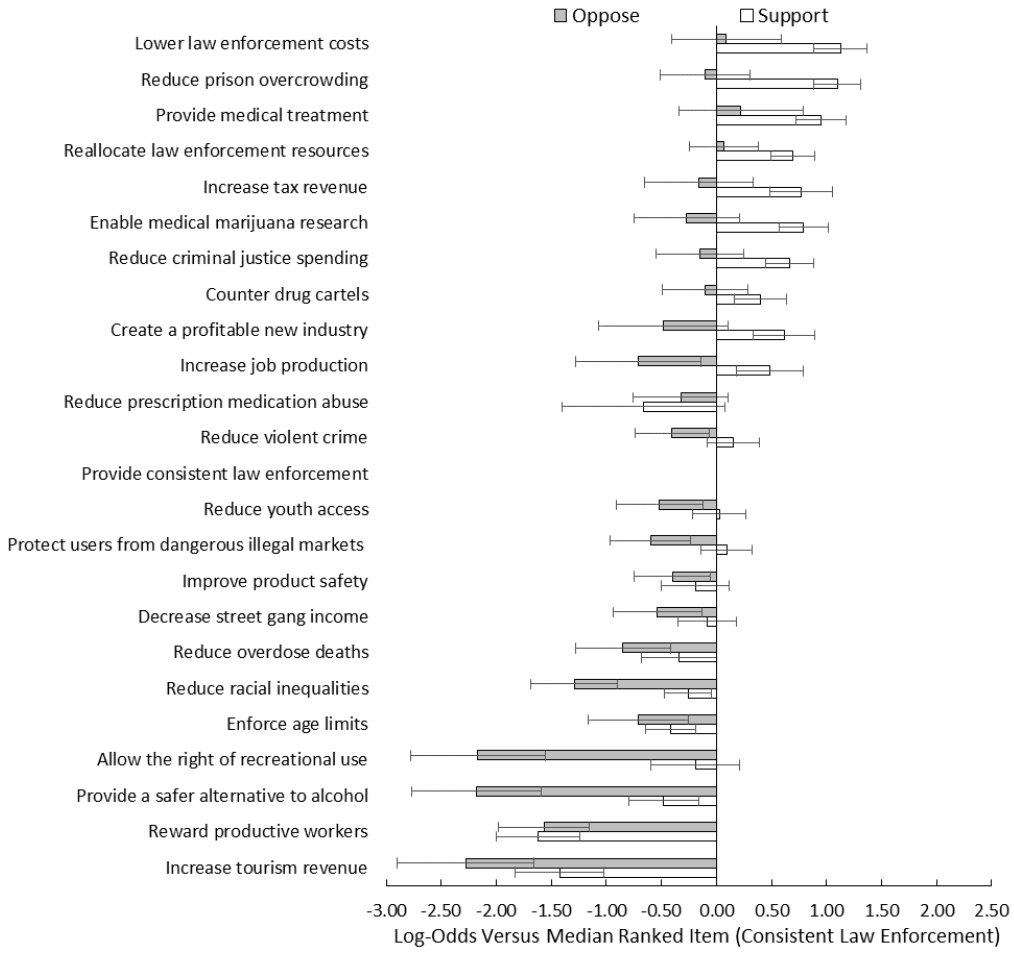

Figure 3. Relative strength of prolegalization arguments by individual-level support or opposition to legalization.

Note: Error bars depict the 95% CI for the estimated log odds of winning for a statement against the median performing statement.

Figure 4 presents parallel information for antilegalization arguments. The data reveal striking patterns of difference in PAS relative to the median by valence of policy preference, as evidenced by greater spread from

the median-ranked argument. This is evident in the figures by examining the difference in log odds between the higher and lower rated arguments.

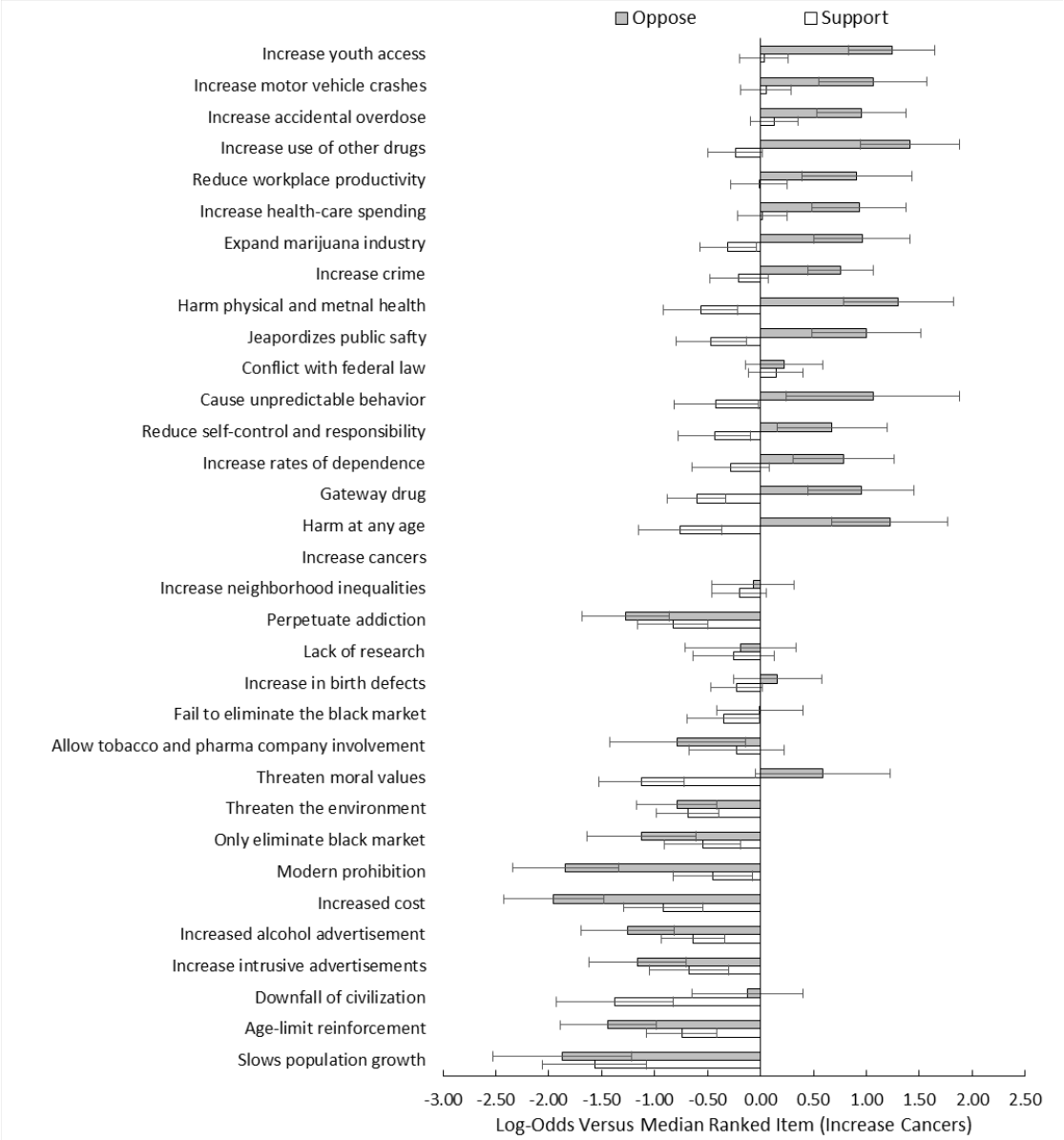

Figure 4. Relative strength of antilegalization arguments by individual-level support or opposition to legalization.

Note: Error bars depict the 95% CI for the estimated log odds of winning for a statement against the median performing statement.

As shown in Figure 4, those opposed to legalization more clearly distinguished between strong and weak arguments than did those in support of the policy. Patterns were similar for prolegalization arguments by individual-level support or opposition to the policy, though differences were clearer for strong arguments.

Discussion

The current study examined whether a survey-embedded wiki platform could provide data about the relative strength of arguments around a dynamic policy issue. We contend that wiki surveys have considerable promise as a method for gauging PAS in dynamic and competitive information environments. We make this claim for several reasons.

The method identified 11 new, user-generated prolegalization arguments and 21 new antilegalization arguments. This is a notable volume because we began the study with a large set of pro- and antilegalization arguments that had appeared in public discourse surrounding the issue. Identifying these new arguments would have taken considerable time via qualitative methods or another concurrent content analysis, as well as additional time to incorporate them in to a subsequent population survey to generate quantitative rankings. Several of these were among the highest rated prolegalization arguments in the study. Previous work found that wiki surveys can quickly identify highly rated ideas when researchers began with a small set of statements that have not been informed by content analyses or public opinion polls (Salganik & Levy, 2015). Our current study began with a large set of arguments based on large-scale analyses of public discourse and opinion. The wiki survey thus passed a conservative test of its ability to rapidly identify new arguments for an evolving policy issue. Combined, these two studies suggest that one could conduct a wiki survey to gauge PAS without taking the time and effort to content analyze arguments in public discourse, though future work should test the extent to which wiki survey results differ depending on the breadth and source of initial arguments.

Although none of the new antilegalization arguments were rated as particularly strong, they nonetheless captured a wider snapshot of the kinds of arguments that citizens may consider when thinking about the issue. This is important for theoretical and practical reasons. Theoretically, the ability to develop robust tests of the ELM and CFT requires a priori knowledge of what arguments are perceived as strong and weak among targeted audiences. Wiki surveys clearly provide useful information in that regard. Practically speaking, public communicators could use this method to identify any oppositional arguments likely to emerge in public debates about contentious policy issues. This knowledge would be useful for developing and disseminating effective countermessages. One could also envision researchers from many fields using this method to inform the design of studies of message effects, public opinion, psychological processes, and/or social trends. Wiki surveys could be used to ensure that studies use language and arguments that characterize current public debate about current social issues when designing messages, scenarios, and/or survey questions. Although we acknowledge that focus groups can produce richer data on new ideas than can a wiki survey, they are time intensive. The wiki survey offers a clear advantage over relying solely on closed-ended surveys (where user-generated items cannot be incorporated) for circumstances where rapid surveillance and assessment is needed.

The wiki survey method produced ratings of argument strength that were similar to more traditional, closed-ended scale-based assessment techniques for prolegalization arguments (providing evidence of predictive validity in one context) although quite different for antilegalization arguments. We are unable to disentangle various potential explanations for the observed differences for antilegalization arguments, which include variation in sample demographics, survey timing (October 2016 vs. April 2016; new arguments may have emerged or become more salient in the months between), and differences in the method itself. Nevertheless, another related wiki survey advantage is that some respondents do not see the full set of arguments a priori—only a randomly chosen subset—which may limit any sources of bias that may be introduced by seeing the entire set. This contrasts with traditional surveys that usually place similar items on one or two pages and ask respondents to rate them simultaneously.

Third, the method offered useful differentiation in PAS among relevant demographic groups, a significant advantage to embedding a wiki platform within a traditional survey. This suggests that wiki surveys can provide useful information to inform the design of strategic messages targeting particular audiences or populations. For instance, knowing which arguments appeal to those already in favor of (or opposed to) a policy could be used to increase political mobilization and policy activism among those groups. Similarly, knowing which arguments resonate among those opposed to (or supporting) the policy could provide clear guidance for strategic efforts to persuade them to support it (or oppose it; Hornik & Woolf, 1999).

The arguments for which this differentiation was clearest depended on the subpopulations being compared. For instance, there was considerable variation in response to antilegalization arguments, but not prolegalization arguments, by a priori policy support. In contrast, there was greater variation in response to prolegalization arguments by state legalization status than there was in response to antilegalization arguments by state legalization status. Our data do not allow us to explore mechanistic explanations for these results, although a few speculative explanations seem plausible. For instance, it may be that differentiation in PAS between those living in states that had legalized recreational marijuana versus those states that had not could have occurred because of experience with some of these arguments (e.g., seeing reduced prison overcrowding, witnessing the benefits of increased tax revenue), focused public debate about the practicalities of implementing such a policy (rather than more abstract moral or ideological concerns that characterized many antilegalization arguments), and/or increased deliberation and scrutiny of the arguments themselves (because surely the law has been a topic of many conversations). The wiki survey revealed several patterns that were not apparent in a traditional, closed-ended survey (from McGinty et al., 2017) and thus may offer a degree of nuance and precision that other methods do not capture as readily.

Researcher Considerations for Using a Survey-Embedded Wiki Platform

We encountered several challenges with the method that are worthy of further scrutiny. First, wiki surveys provide a relative assessment of strength, not an absolute one. One can imagine topics for which there are no strong proarguments, in absolute terms (e.g., brushing one's teeth with cola). The wiki survey approach, in isolation, would not be able to identify such a topic. Thus, the method may work best for public issues for which the researcher is confident, based on some a priori assessment of public opinion, that reasonable anti- and proarguments exist in public discourse. Second, some respondents were confused

about the process (e.g., "the never-ending survey") despite very clear instructions. Studies may benefit from having respondents practice the method on an unrelated topic before proceeding to the main study.

Second, future wiki surveys could also impose a time limit on a wiki survey page before moving to the next page, as some respondents encountered a very large set of item comparisons. Although our analytic strategy of clustering on the individual identifier corrects for any skewing of results that high-volume raters could have on results, imposing some limitations on the volume of individual ratings could be useful in future work.

Third, the user-generated ideas require screening before incorporation into the pool of available arguments. The majority of user-generated ideas were commentary on existing arguments or duplicative of existing ones, not new information. Novel arguments (before screening by the research team) were often unduly complex or double-barreled. Thus, there is a nontrivial burden on the research team to screen and vet user-generated commentary, and this process must happen quickly to incorporate this information into the larger pool of arguments.

Fourth, wiki surveys may work best when data are collected over several weeks, giving sufficient time to allow novel arguments to accumulate enough rankings to produce stable estimates. Finally, the nature of the data and their structure requires a more complex analytic approach than is typically required in the analysis of survey data (e.g., mean ratings; clustering to account for respondent variation in the number of pairwise comparisons).

PAS and Actual Message Persuasiveness

Recent field debates have raised questions about the predictive validity of PME, a concept closely related to PAS (Nabi, 2018; O'Keefe, 2018). Although this debate is far from settled, and many studies have found that messages high in PAS are indeed more persuasive (Cappella, 2018; Zhao et al., 2011), these concerns underscore the need to assess the predictive validity of the wiki survey approach in future work. Although our study directly compared wiki survey results with those obtained by traditional, closed-ended scale measures (McGinty et al., 2017), we do not provide a direct test of whether wiki survey rankings of PAS predict which arguments are more persuasive in changing policy support. Wiki survey rankings differed from traditional survey methods for antilegalization arguments, raising questions about which method is better at identifying arguments that change attitudes or intentions. Thus, a randomized survey experiment testing whether wiki-survey-identified strong arguments outperform wiki-survey-identified moderate and weak arguments or strong arguments identified in a traditional survey format is an important next step.

Other Study Limitations and Future Directions

This current study is limited by its focus on a single topic during a limited time period; additional studies in other contexts are needed to further clarify the benefits and trade-offs of the wiki survey method. The study's response rate was low (10%), limiting the generalizability of the substantive findings about recreational marijuana legalization. However, we view this work as a proof of concept and invitation for future work with the method. Wiki surveys could be incorporated into Web-based surveys that employ

probability sampling (e.g., GfK's KnowledgePanel; NORC's AmeriSpeak Panel) and offer larger incentives to increase response rates.

We nevertheless believe these data and analysis support the promise of broader use and continued refinement of the method in message effects and public opinion research, as the method's potential benefits (i.e., rapid identification and incorporation of novel arguments without having to conduct a large-scale scan of the public information environment) are substantial.

References

Atkin, C. K., & Freimuth, V. S. (2001). Formative evaluation research in campaign design. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* (3rd. ed., pp. 125–146). Thousand Oaks, CA: SAGE Publications.

Baek, Y. M., Cappella, J. N., & Bindman, A. (2011). Automating content analysis of open-ended responses: Wordscores and affective intonation. *Communication Methods and Measures, 5*, 275–296. https://doi:10.1080/19312458.2011.624489

Cappella, J. N. (2018). Perceived message effectiveness meets the requirements of a reliable, valid, and efficient measure of persuasiveness. *Journal of Communication, 68*, 994–997. https://doi:10.1093/joc/jqy044

Carpenter, C. J. (2015). A meta-analysis of the ELM's argument quality × processing type predictions. *Human Communication Research, 41*, 501–534. https://doi:10.1111/hcre.12054

Chong, D., & Druckman, J. N. (2007a). A theory of framing and opinion formation in competitive elite environments. *Journal of Communication, 57*, 99–118. https://doi:10.1111/j.1460-2466.2006.00331.x

Chong, D., & Druckman, J. N. (2007b). Framing public opinion in competitive democracies. *American Political Science Review, 101*, 637–655. https://doi:10.1017/S0003055407070554

Chong, D., & Druckman, J. N. (2013). Counterframing effects. *The Journal of Politics, 75*, 1–16. https://doi:10.1017/S0022381612000837

Dillard, J. P., & Ye, S. (2008). The perceived effectiveness of persuasive messages: Questions of structure, referent, and bias. *Journal of Health Communication, 13*, 149–168. https://doi:10.1080/10810730701854060

Else, G. F. (1957). *Aristotle's poetics: The argument*. Cambridge, MA: Harvard University Press.

Fishbein, M., & Ajzen, I. (2009). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Geiger, A. (2016, October 12). *Support for marijuana legalization continues to rise*. Retrieved from http://www.pewresearch.org/fact-tank/2016/10/12/support-for-marijuana-legalization-continues-to-rise/

Gollust, S., Niederdeppe, J., & Barry, C. (2013). Framing the consequences of childhood obesity to increase support for obesity prevention policy. *American Journal of Public Health, 103,* e96–e102. https://doi:10.2105/AJPH.2013.301271

Hornik, R., & Woolf, K. D. (1999). Using cross-sectional surveys to plan message strategies. *Social Marketing Quarterly, 5*, 34–41. https://doi:10.1080/15245004.1999.9961044

McGinty, E. E., Niederdeppe, J., Heley, K., & Barry, C. L. (2017). Public perceptions of arguments supporting and opposing recreational marijuana legalization. *Preventive Medicine, 99*, 80–86. https://doi:10.1016/j.ypmed.2017.01.024

McGinty, E. E., Samples, H., Bandara, S., Saloner, B., Bachhuber, M., & Barry, C. L. (2016). The emerging public discourse on state legalization of marijuana for recreational use in the U.S.: Analysis of news media coverage, 2010–2014. *Preventive Medicine, 90*, 114–120. https://doi:10.1016/j.ypmed.2016.06.040

Nabi, R. L. (2018). On the value of perceived message effectiveness as a predictor of actual message effectiveness: An introduction. *Journal of Communication, 68*, 988–989. https://doi:10.1093/joc/jqy048

O'Keefe, D. J. (2018). Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication, 68*, 120–142. https://doi:10.1093/joc/jqx009

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123–205. https://doi:10.1016/S0065-2601(08)60214-2

Petty, R. E., & Wegener, D T. (1998). Attitude change. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 323–390). New York, NY: McGraw-Hill.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Salganik, M. J., & Levy, K. E. C. (2015). Wiki surveys: Open and quantifiable social data collection. *PLOS ONE 10*(5), e0123483. https://doi:10.1371/journal.pone.0123483

Tarico, V. S., Van Velzen, D. R., & Altmaier, E. M. (1986). Comparison of thought-listing rating methods. *Journal of Counseling Psychology, 33*, 81–83. https://doi:10.1037/0022-0167.33.1.81

Turner, H., & Firth, D. (2012). Bradley-Terry models in R: The BradleyTerry2 package. *Journal of Statistical Software, 48*(9), 1–21.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge, UK: Cambridge University Press.

Yzer, M., LoRusso, S., & Nagler, R. H. (2015). On the conceptual ambiguity surrounding perceived message effectiveness. *Health Communication, 30*, 125–134. https://doi:10.1080/10410236.2014.974131

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software, 11*(10), 1–17. https://doi:10.18637/jss.v011.i10

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software, 16*(9), 1–16. https://doi:10.18637/jss.v016.i09

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*, 48–75. https://doi:10.1080/19312458.2010.547822